1943

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

H. J. C.

SECOND SERIES, VOL. 44

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

T
boo
not
mer
inte
a Pr
exa
unk

is sc

does
In
ellip
forn
mad
L
*u,
use
p, n
who
L
In
In

1
Lonc
para
diffe
effor
Lonc
2
D. H
1912,
3 T
stan
woul

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

ON HODGE'S THEORY OF HARMONIC INTEGRALS

BY HERMANN WEYL

(Received June 4, 1942)

The attempt which W. V. D. Hodge made in Chapter III of his beautiful book[1] to establish *the existence of harmonic integrals with preassigned periods* has not been entirely successful because the proof is partly based on a false statement (p. 136) concerning the behavior of the solution of a non-homogeneous integral equation when the spectrum parameter approaches an eigen value. In a Princeton seminar on the subject, H. F. Bohnenblust pointed out that counter examples are readily available even for linear equations with a finite number of unknowns. For instance the equation $\lambda x + Ax = c$ with

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is solvable for $\lambda = 0$ (x_1 arbitrary, $x_2 = 1$) and yet the solution for $\lambda \neq 0$,

$$x_1 = 1/\lambda, \qquad x_2 = 0$$

does not tend to a limit with $\lambda \to 0$.

In his book Hodge uses the *parametrix method* first developed for a single elliptic differential equation by E. E. Levi and D. Hilbert.[2] Building on the formal foundations laid by Hodge, I will show here how the argument can be made conclusive. Hilbert's procedure served me as a model.

Let n be the dimensionality of our Riemannian manifold. I denote by $^*u$, Du the dual form and the derivative of any (linear differential) form u and use the abbreviation Δ for the operator D^*D.[3] For two forms u, v of rank p, $n - p$ respectively (v, u) designates the integral of the product $v \cdot u$ over the whole manifold. $(^*u, u)$ is positive unless $u = 0$. An immediate consequence is

LEMMA I. *$\Delta u = 0$ implies $Du = 0$.*

Indeed $(D^*Du, u) = 0$ leads to $(^*Du, Du) = 0$, hence $Du = 0$.

In the following, f, u, φ, η are forms of rank p and g, v, ψ, ϑ forms of rank $n - p$.

---

[1] *The Theory and Applications of Harmonic Integrals*, Cambridge, 1941. See also Proc. London Math. Soc. (2) **41**, 1936, pp. 483–496 where Hodge ascribes the idea of using Hilbert's parametrix method to H. Kneser. I find it hard to judge whether a previous proof along different lines (Proc. London Math. Soc. (2) **38**, 1933, p. 72) is complete, or rather how much effort is needed to make it complete. For the Euclidean case, see W. V. D. Hodge, Proc. London Math. Soc. (2) **36**, 1932, p. 257, and H. Weyl, Duke Math. Jour. **7**, 1940, pp. 411–444.

[2] E. E. Levi, Memorie della Società italiana delle Scienze, Ser. 3[a], Tom. 16, 1909; D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, Leipzig, 1912, pp. 223–231.

[3] The author intended $*u$, the printer evidently disliked it and replaced it by $^*u$. If a standardized notation in the theory of linear differential forms is adopted the author would recommend to follow him and not the printer.

The rank p is fixed; no induction with respect to p takes place. The goal is to prove the following

THEOREM I. *For any given null form* g, $g \sim 0$, *the equation*

$$\Delta u = g \tag{1}$$

has a solution u.

I copy Hodge's two basic formulas (3) and (4) on pp. 132, 133 of his book, replacing $p - 1$ by p and using the abbreviation $1/\gamma = (-1)^{np}(n - 2)\alpha_n$. Let K be the operator with the kernel $\gamma \cdot K_p(x, y)$ which carries any form $u(x)$ into $\gamma \int K_p(x, y) \cdot u(y)$, and K' its transpose. The "parametrix" operators Q, P with the kernels $\gamma \cdot \omega_p(x, y)$ and $\gamma \cdot \omega_{p-1}(x, y)$ are symmetric,

$$(Qv, g) = (Qg, v).$$

Finally, I set $DPD = \Pi$. Hodge's formulas read

$$Ku - u = Q\Delta u + (-1)^n \Pi^* u, \tag{2}$$

$$K'v - v = \Delta Qv + (-1)^{n*}\Pi v. \tag{2'}$$

The solutions of the equations

$$Ku - u = 0, \qquad K'v - v = 0$$

will be called the eigen forms of the kernels K and K' (*scilicet* "for the eigen value 1"). We try to solve our problem by means of the non-homogeneous integral equation suggested by (2),

$$Ku - u = Qg. \tag{E}$$

It is essential to study this equation not only for null forms but in a wider set $\mathcal{G}$; the success of the method depends on the proper choice of that linear space $\mathcal{G}$. Here is my definition:

g belongs to $\mathcal{G}$ whenever PDg is closed,

$$DPDg = \Pi g = 0.$$

Every form of the type

$$f = \Pi v \qquad (v \text{ arbitrary})$$

is said to belong to $\mathcal{F}$. Evidently $\mathcal{G}$ contains all *closed* forms g whereas all elements f of $\mathcal{F}$ are *null* forms. $\mathcal{F}$ and $\mathcal{G}$ are orthogonal:

LEMMA II. $(g, f) = 0$ *for* $g \in \mathcal{G}, f \in \mathcal{F}$.

Indeed, if PDg is closed, then

$$(PDg, Dv) = 0 = (Dg, PDv),$$

an equation which may at once be changed into

$$(g, DPDv) = 0.$$

I take over Hodge's Lemma I on p. 142:

LEMMA III. *If ψ is any eigen form of K' then $Q\psi$ is closed.*

For the sake of completeness I repeat the simple proof. Equation (2′) yields for $\xi = Q\psi$:

$$\Delta\xi = (-1)^{n-1} {}^*\Pi\psi, \tag{3}$$

hence $D^*\Delta\xi = D^*D^*D\xi = 0$ and then by double application of Lemma I,[4]

$$D^*D\xi = 0, \qquad D\xi = 0.$$

Incidentally we learn from (3) and the intermediate equation $\Delta\xi = 0$ that $\Pi\psi = 0$, or that *the eigen forms ψ of K' lie in $\mathcal{G}$.*

We analyze the eigen forms of K and K' as follows. Within the linear space of all eigen forms $\bar{\varphi}$ of K we consider the subspace $\mathfrak{f}$ of the *closed* eigen forms φ and choose our basis

$$\varphi_1, \cdots, \varphi_l, \qquad \bar{\varphi}_1, \cdots, \bar{\varphi}_m$$

for all eigen forms accordingly, i.e. $\varphi_1, \cdots, \varphi_l$ span $\mathfrak{f}$. Equation (2) yields

$$Q\Delta\bar{\varphi} = (-1)^{n-1}\Pi^*\bar{\varphi}. \tag{4}$$

This proves on the one hand that each closed eigen form φ of K satisfies the condition $\Pi^*\varphi = 0$,

LEMMA IV. ${}^*\varphi \in \mathcal{G}$ *for every* $\varphi \in \mathfrak{f}$.

It shows on the other hand that $\bar{\psi} = \Delta\bar{\varphi}$ satisfies the conditions

$$\Delta Q\bar{\psi} = 0, \qquad \Pi\bar{\psi} = 0$$

because the operators $\Delta\Pi$ and $\Pi\Delta$ annihilate. It then follows from (2′) that $\bar{\psi}$ is an eigen form of K'. The m forms $D\bar{\varphi}_1, \cdots, D\bar{\varphi}_m$ are linearly independent by construction, and hence by Lemma I the same is true for the forms

$$\bar{\psi}_1 = \Delta\bar{\varphi}_1, \cdots, \qquad \bar{\psi}_m = \Delta\bar{\varphi}_m.$$

The transposed kernel K' has the same number $l + m$ of linearly independent eigen forms as K. We determine a basis

$$\bar{\psi}_1, \cdots, \bar{\psi}_m; \qquad \psi_1, \cdots, \psi_l \tag{5}$$

of which the $\bar{\psi}$'s are a part.

The integral equation (E) is solvable if and only if

$$(Qg, \psi) = 0 = (g, Q\psi)$$

for every eigen form ψ of K', or with the notation $\xi = Q\psi$, if

$$(g, \xi) = 0. \tag{6}$$

[4] One differentiation may be saved here by applying the formula $(Ds, Dt) = 0$ holding for any two forms s, t with continuous first derivatives of rank $p - 1$ and $n - p - 1$ (see Weyl, l.c.[1], p. 426) to $s = PD\psi$ and $t = {}^*D\xi$ with the result $(\Pi\psi, \Delta\xi) = 0 = ({}^*\Delta\xi, \Delta\xi)$ whence $\Delta\xi = 0 = \Pi\psi$.

Let us say that ψ is *of the first kind* when $\xi = Q\psi \in \mathfrak{F}$. The forms $\bar{\psi}_1, \cdots, \bar{\psi}_m$ are of the first kind, on account of the equation (4). We choose our basis (5) so that

$$\bar{\psi}_1, \cdots, \bar{\psi}_m; \qquad \psi_1, \cdots, \psi_\nu$$

span the linear manifold of all eigen forms of K' of the first kind. By Lemma II the relation (6) holds good for *any* $g \in \mathfrak{G}$ in case ψ is of the first kind, and thus the $m + l$ conditions (6) reduce to the last $l - \nu$ of them,

$$(g, Q\psi_{\nu+1}) = 0, \cdots, \qquad (g, Q\psi_l) = 0. \tag{7}$$

Let $\mathfrak{G}_1$ denote the set of those forms $g \in \mathfrak{G}$ which satisfy the conditions (7). *We have found that under the assumption $g \in \mathfrak{G}_1$ the integral equation* (E) *has a solution u.*

For this solution u we obtain from (2):

$$Q(g - \Delta u) = (-1)^n \cdot \Pi^* u, \tag{8}$$

hence $\Delta Q(g - \Delta u) = 0$. Combining this with $\Pi(g - \Delta u) = 0$ and applying (2′) to $v = g - \Delta u$ one finds

$$g - \Delta u = \psi = \bar{c}_1\bar{\psi}_1 + \cdots + \bar{c}_m\bar{\psi}_m + c_1\psi_1 + \cdots + c_l\psi_l$$

to be an eigen form of K'. More precisely, because of (8), $Q\psi \in \mathfrak{F}$, ψ is an eigen form of the first kind, which forces $c_{\nu+1}, \cdots, c_l$ to vanish. Writing u for $u + \bar{c}_1\bar{\varphi}_1 + \cdots + \bar{c}_m\bar{\varphi}_m$ we arrive at the following

Intermediary Proposition: *For any $g \in \mathfrak{G}_1$ there exists a form u and ν constants $c_1, \cdots, c_\nu$ such that*

$$g - \Delta u = c_1\psi_1 + \cdots + c_\nu\psi_\nu . \tag{9}$$

We know from Lemma IV that the dual form $^*\varphi$ of any element φ of $\mathfrak{f}$ lies in $\mathfrak{G}$. That subspace of $\mathfrak{f}$ the elements φ of which satisfy the conditions

$$(^*\varphi, Q\psi_{\nu+1}) = 0, \ \cdots, \ (^*\varphi, Q\psi_l) = 0$$

is of a dimensionality $\mu \geqq \nu$. Let the basis $\varphi_1, \cdots, \varphi_l$ of $\mathfrak{f}$ be so chosen that $\varphi_1, \cdots, \varphi_\mu$ span this subspace. From (9) we obtain for the ν unknowns c_β the μ linear equations

$$\sum\nolimits_\beta H_{\alpha\beta} \cdot c_\beta = (g, \varphi_\alpha) \qquad \begin{pmatrix} \alpha = 1, \cdots, \mu; \\ \beta = 1, \cdots, \nu \end{pmatrix} \tag{10}$$

where

$$H_{\alpha\beta} = (\psi_\beta, \varphi_\alpha).$$

I maintain:

Lemma v. *$\| H_{\alpha\beta} \|$ is a non-singular square matrix.*

Once this is established we have reached the goal. For then the ν conditions

$$(g, \varphi_\alpha) = 0 \qquad (\alpha = 1, \cdots, \nu)$$

imply $c_\alpha = 0$ whereby (9) reduces to $g - \Delta u = 0$. In other words, if $g \in \mathfrak{G}$ satisfies the relations

$$(11) \quad (g, Q\psi_{\nu+1}) = 0, \ \cdots, \ (g, Q\psi_l) = 0; \qquad (g, \varphi_1) = 0, \ \cdots, \ (g, \varphi_\nu) = 0$$

then the equation (1) is solvable. A null form g fulfills all our requirements, because the φ_i and $Q\psi_i$ are closed, the first by construction, the others by Lemma III.

Proof of Lemma v. We have found the equations (10) to be solvable if $g \in \mathfrak{G}_1$. For

$$(12) \qquad \varphi = a_1\varphi_1 + \cdots + a_\mu\varphi_\mu$$

the integral $({}^*\varphi, \varphi)$ is a positive definite quadratic form of $a_1, \cdots, a_\mu$. Hence we can determine the coefficients a_i in (12) so as to assign arbitrary values b_α to the integrals

$$({}^*\varphi, \varphi_\alpha) \qquad (\alpha = 1, \cdots, \mu).$$

But $g = {}^*\varphi \in \mathfrak{G}_1$. Hence we see that the equations

$$\sum_\beta H_{\alpha\beta} c_\beta = b_\alpha \qquad \begin{pmatrix} \alpha = 1, \cdots, \mu; \\ \beta = 1, \cdots, \nu \end{pmatrix}$$

have a solution c_β for arbitrary b_α. In view of $\mu \geqq \nu$ this statement is equivalent to our lemma.

In proving Theorem I we actually showed that the equation $\Delta u = g$ is solvable if $g \in \mathfrak{G}$ satisfies the conditions (11). Hence each such g is a null form, and the linear space $\mathfrak{G}$ is of finite dimensionality $\leqq l$ modulo the space of null forms. As $\mathfrak{G}$ contains all closed forms of rank $n - p$, we find *a fortiori* that the number R'_{n-p} of linearly independent closed forms of rank $n - p$ modulo null is finite and $\leqq l$. The conditions (11) are of the type $(g, f) = 0$ where f runs over certain specified closed forms of rank p. Consider the "inner product" (g, f) of any two closed forms g, f of rank $n - p$ and p respectively; the factors matter only modulo null. Our proof implies this further fact:

Theorem ii. *If the inner product (g, f) vanishes for a given closed g and all closed f, then $g \sim 0$.*

It is of course also true that the product cannot vanish for a given closed f and all closed g unless $f \sim 0$. Both facts together give the duality law

$$(13) \qquad R'_{n-p} = R'_p .$$

Theorem II has nothing to do with any Riemannian metric. de Rham's second theorem follows at once from it by means of the expression of the product (g, f) in terms of the periods of g and f (Hodge, p. 85, last line), but it is essentially simpler since it deals with closed forms only, and not with forms and cycles. Its proof on an arbitrary manifold should be correspondingly easier.

The following proposition is equivalent to Theorem I for the rank $p - 1$ instead of p:

THEOREM III. *For any form f there exists a uniquely determined $\eta \sim f$ such that $^*\eta$ is closed. If f be closed, then η is harmonic.*

Indeed, set $f = Dt + \eta$, t being of rank $p - 1$. The requirement $D^*\eta = 0$ leads to the equation $D^*Dt = D^*f$ which is solvable by Theorem I.

The new proposition shows at once that for any rank p the space of closed forms modulo null may be identified with the space of harmonic forms. This makes the equation (13) particularly lucid because $^*u$ is harmonic if u is and vice versa. The same proposition provides another proof for Theorem II, because one has merely to substitute $^*\eta$ for ϑ in order to see that the vanishing of the inner product (η, ϑ) of a fixed harmonic form η with every harmonic ϑ implies $\eta = 0$. The observation (Hodge, p. 139) that on account of (2) the harmonic p-forms are eigen forms of K again proves the inequality $R'_p \leqq l$.

The link with the homology theory of cycles is established by de Rham's first theorem stating that a p-cycle C is homologous zero if the integral of every closed p-form f over C vanishes.

INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J.

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

MOUTARD-ČECH HYPERQUADRICS ASSOCIATED WITH A POINT OF A HYPERSURFACE

BY BUCHIN-SU

(Received May 22, 1942)

The purpose of this paper is to give a new proof of the theorem of Čech[1] concerning the locus of the quadrics of Čech of sections produced by certain spaces and to generalize, as a consequence of the equation to a Moutard-Čech hyperquadric, the notion of Moutard pencils[2] to the theory of hypersurfaces.

Let O be a generic point of a hypersurface V_n in a projective space S_{n+1}. If O is taken for the origin of coordinates with the tangent hyperplane of V_n at O in non-homogeneous coordinates $X^1, \cdots, X^{n+1}$ as the coordinate hyperplane $X^{n+1} = 0$, then the expansion of V_n at O is of the form

$$X^{n+1} = \tfrac{1}{2}H_{\sigma\tau}X^\sigma X^\tau + \tfrac{1}{3}K_{\sigma\tau\rho}X^\sigma X^\tau X^\rho + \tfrac{1}{12}H_{\sigma\tau\rho u}X^\sigma X^\tau X^\rho X^u + \cdots, \tag{1}$$

where we have made use of the convention that when the same index appears as a subscript and superscript in a term this term stands for the sum of the terms obtained by giving the index each of its n values $1, 2, \cdots, n$. Without loss of generality we may further assume that the coefficients in (1) are symmetrical in the subscripts.

We shall calculate the fundamental quantities $a_{ij}(i, j = 1, \cdots, n)$ of the fundamental form F_2 of Fubini.[3] Putting

$$b_{\sigma\tau}\, dX^\sigma dX^\tau = \left(X \frac{\partial X}{\partial X^1} \cdots \frac{\partial X}{\partial X^n}\, d^2X\right)$$

for homogeneous coordinates with $X^0 = 1$, and noticing that

$$\frac{\partial X^i}{\partial X^j} = \delta^i_j \qquad (i, j = 1, \cdots, n),$$

we have

$$b_{rs} = \frac{\partial^2 X^{n+1}}{\partial X^r \partial X^s} = H_{rs} + 2K_{rs\rho}X^\rho + H_{rs\rho\tau}X^\rho X^\tau + (3), \tag{2}$$

where (m) denotes all of the terms of order $\geqq m$ in $X^1, \cdots, X^n$. The discriminant B of the above form is consequently found to be

$$B = H\{1 + 2H^{\sigma\tau}K_{\sigma\tau\rho}X^\rho + L_{\sigma\tau}X^\sigma X^\tau + (3)\}, \tag{3}$$

[1] Cf. G. Fubini and E. Čech, *Geometria proiettiva differenziale*, vol. II, Bologna (1927), p. 618; in the sequel referred to as *G. P. D.*

[2] For the definition of a Moutard pencil in S_3 see B. Su and A. Ichida, *On certain cones connected with a surface in the affine space*, Japanese Journal of Mathematics, 10 (1933), pp. 209–216.

[3] Cf. *G. P. D.*, p. 608.

where

$$L_{vw} = H^{\sigma\tau}H_{\sigma\tau vw} + \tfrac{1}{2}H^{\sigma u,\tau\rho}(K_{\sigma\tau v}K_{u\rho w} - K_{u\tau v}K_{\sigma\rho w} + K_{\sigma\tau w}K_{u\rho v} - K_{u\tau w}K_{\sigma\rho v}) \tag{4}$$

and H is the determinant of H_{rs}, H^{ij} is the algebraic complementary of H_{ij} in H, divided by H and, finally, $H^{ij,kl} = 0$ if $i = j$ or $k = l$, and otherwise $H^{ij,kl}$ is the minor formed by striking out the rows and columns in H which contain H_{ik}, H_{il}, H_{jk}, H_{jl}, also divided by H, the sign being $+1$ or -1 according as the permutations i, j and the remaining rows in order, and k, l and the remaining columns in the order $1, 2, \cdots, n$ are of the same or opposite parity.

From (3) it follows that

$$\begin{aligned} B^{-1/(n+2)} = H^{-1/(n+2)} \Bigg\{ 1 &- \frac{2}{n+2} H^{\sigma\tau}K_{\sigma\tau\rho}X^{\rho} \\ &+ \left(\frac{2(n+3)}{(n+2)^2} H^{\sigma\tau}H^{\rho u}K_{\sigma\tau v}K_{\rho uw} - \frac{1}{n+2}L_{vw}\right)X^{v}X^{w} + (3)\Bigg\}, \end{aligned} \tag{5}$$

so that

$$\begin{aligned} a_{rs} &= B^{-1/(n+2)}b_{rs} \\ &= H^{-1/(n+2)}\left\{H_{rs} + 2\left(K_{rs\rho} - \frac{1}{n+2}H_{rs}H^{\sigma\tau}K_{\sigma\tau\rho}\right)X^{\rho} + (2)\right\}. \end{aligned} \tag{6}$$

Thus we obtain *at O*

$$\left\{\begin{aligned} a_{rs} &= H^{-1/(n+2)}H_{rs} && (r, s = 1, \cdots, n), \\ \frac{\partial a_{rs}}{\partial X^{t}} &= 2H^{-1/(n+2)}\left(K_{rst} - \frac{1}{n+2}H_{rs}H^{\sigma\tau}K_{\sigma\tau t}\right) && (r, s, t, = 1, \cdots, n), \\ a^{rs} &= H^{1/(n+2)}H^{rs}, \\ \Gamma_{rs,t} &= \frac{1}{2}\left(\frac{\partial a_{rt}}{\partial X^{s}} + \frac{\partial a_{st}}{\partial X^{r}} - \frac{\partial a_{rs}}{\partial X^{t}}\right) \\ &= H^{-1/(n+2)}\Bigg\{K_{rst} - \frac{1}{n+2}H_{rt}H^{\sigma\tau}K_{\sigma\tau s} \\ &\qquad - \frac{1}{n+2}H_{st}H^{\sigma\tau}K_{\sigma\tau r} + \frac{1}{n+2}H_{rs}H^{\sigma\tau}K_{\sigma\tau t}\Bigg\}, \\ \Gamma^{l}_{rs} &= H^{l\sigma}K_{rs\sigma} - \frac{1}{n+2}H^{\sigma\tau}K_{\sigma\tau s}\delta^{l}_{r} \\ &\qquad - \frac{1}{n+2}H^{\sigma\tau}K_{\sigma\tau r}\delta^{l}_{s} + \frac{1}{n+2}H^{l\rho}H_{rs}H^{\sigma\tau}K_{\sigma\tau\rho}. \end{aligned}\right. \tag{7}$$

The covariant derivatives of X^{i} are

$$X^{i}_{rs} = \frac{\partial^2 X^{i}}{\partial X^{r}\partial X^{s}} - \Gamma^{l}_{rs}\frac{\partial X^{i}}{\partial X^{l}} \qquad (i = 0, 1, \cdots, n+1).$$

At O we obtained from (7), X^0 being 1,

$$
\begin{cases} X^0_{rs} = 0, \\ X^i_{rs} = -\Gamma^i_{rs} \qquad (i = 1, \cdots, n), \\ X^{n+1}_{rs} = H_{rs}. \end{cases} \tag{8}
$$

Introduce, as usual, the vector $\mathfrak{X}$ of components[4]

$$
\mathfrak{X}^i = \frac{1}{n}\Delta_2 X^i = \frac{1}{n} a^{\sigma\tau} X^i_{\sigma\tau} \qquad (i = 0, 1, \cdots, n+1) \tag{9}
$$

and calculate the values at O by means of (7) and (8). The result of this computation is as follows:

$$
\begin{cases} \mathfrak{X}^0 = 0, \\ \mathfrak{X}^i = -\dfrac{2}{n+2} H^{1/(n+2)} H^{\sigma\tau} K_{\sigma\tau\rho} H^{i\rho} \quad (i = 1, \cdots, n), \\ \mathfrak{X}^{n+1} = H^{1/(n+2)}. \end{cases} \tag{10}
$$

We are in a position to consider the corresponding tangential coordinates $\xi^i (i = 0, 1, \cdots, n+1)$ of Čech given by the minors of order n in the following matrix[5]:

$$
\left\| \begin{matrix} 1 & X^1 & X^2 & \cdots & X^n & X^{n+1} \\ 0 & 1 & 0 & \cdots & 0 & \dfrac{\partial X^{n+1}}{\partial X^1} \\ 0 & 0 & 1 & \cdots & 0 & \dfrac{\partial X^{n+1}}{\partial X^2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & \dfrac{\partial X^{n+1}}{\partial X^n} \end{matrix} \right\| : B^{1/(n+2)},
$$

or, more precisely,

$$
\begin{cases} \xi^i = B^{-1/(n+2)} \dfrac{\partial X^{n+1}}{\partial X^i} \qquad (i = 1, \cdots, n), \\ \xi^{n+1} = B^{-1/(n+2)}. \end{cases} \tag{11}
$$

A reference to (1) and (5) gives

$$
\xi^i = H^{-1/(n+2)} \left\{ H_{i\rho} X^\rho + \left(K_{ivw} - \frac{2}{n+2} H_{iv} H^{\sigma\tau} K_{\sigma\tau w} \right) X^v X^w + (3) \right\} \tag{12}
$$
$$
(i = 1, \cdots, n).
$$

[4] Cf. *G. P. D.*, p. 611. For clearness of notation the X there utilized is now replaced by $\mathfrak{X}$.

[5] Cf. *G. P. D.*, p. 609.

so that at O

$$\frac{\partial \xi^i}{\partial X^l} = H^{-1/(n+2)} H_{il} \qquad (i, l = 1, \cdots, n),$$

$$\frac{\partial^2 \xi^i}{\partial X^r \partial X^s} = 2H^{-1/(n+2)} \left(K_{irs} - \frac{1}{n+2} H_{ir} H^{\sigma\tau} K_{\sigma\tau s} - \frac{1}{n+2} H_{is} H^{\sigma\tau} K_{\sigma\tau r} \right) \qquad (r, s = 1, \cdots, n)$$

and therefore

$$\xi^i_{rs} = H^{-1/(n+2)} \left\{ K_{irs} - \frac{1}{n+2} H^{\sigma\tau} K_{\sigma\tau s} H_{ir} - \frac{1}{n+2} H^{\sigma\tau} K_{\sigma\tau r} H_{is} - \frac{1}{n+2} H^{\sigma\tau} K_{\sigma\tau i} H_{rs} \right\} \qquad (i = 1, \cdots, n). \tag{13}$$

If we define, as the dual of $\mathfrak{X}$, the vector Ξ of components

$$\Xi^j = \frac{1}{n} \Delta_2 \xi^j = \frac{1}{n} a^{rs} \xi^j_{rs} \qquad (j = 0, 1, \cdots, n+1), \tag{14}$$

we have at 0

$$\begin{aligned} \Xi^i &= 0 \qquad (i = 1, \cdots, n), \\ \Xi^{n+1} &= \frac{4}{n(n+2)} \left\{ \frac{2n+3}{n+2} H^{\sigma\tau} H^{\rho u} H^{rs} K_{\sigma\tau r} K_{\rho u s} - \tfrac{1}{2} H^{rs} L_{rs} \right\}, \end{aligned} \tag{15}$$

whence

$$S\mathfrak{X}\Xi = \frac{4}{n(n+2)} H^{1/(n+2)} \left\{ \frac{2n+3}{n+2} H^{\sigma\tau} H^{\rho u} H^{rs} K_{\sigma\tau r} K_{\rho u s} - \tfrac{1}{2} H^{rs} L_{rs} \right\}. \tag{16}$$

The quadric of Čech of V_n at O is given by the equation[6]

$$2\lambda\mu + (\mu)^2 S\mathfrak{X}\Xi - a_{\sigma\tau} \nu^\sigma \nu^\tau = 0 \tag{17}$$

where the local coordinates $\lambda, \nu^1, \cdots, \nu^n, \mu$ of a point are related to the coordinates $X^1, \cdots, X^{n+1}$ by the following equations:

$$1 : X^1 : X^2 : \cdots : X^n : X^{n+1} = \lambda : \mu\mathfrak{X}^1 + \nu^1 : \mu\mathfrak{X}^2 + \nu^2 : \cdots : \mu\mathfrak{X}^n + \nu^n : \mu\mathfrak{X}^{n+1}. \tag{18}$$

Substituting (10) and (16) into (17) and reducing by (18), we have the equation to the quadric of Čech of V_n at O:

$$X^{n+1} - \tfrac{1}{2} H_{\sigma\tau} X^\sigma X^\tau - \frac{2}{n+2} H^{\rho u} K_{\rho u \sigma} X^\sigma X^{n+1} + \left\{ \frac{2(n+3)}{n(n+2)^2} H^{\sigma\tau} H^{\rho u} H^{rs} K_{\sigma\tau r} K_{\rho u s} - \frac{1}{n(n+2)} H^{rs} L_{rs} \right\} (X^{n+1})^2 = 0. \tag{19}$$

[6] Cf. *G. P. D.*, p. 616.

From the last equation we can easily demonstrate some known results. In fact, consider a hyperquadric Q which has at O a contact of the second order with V_n. Since the equation of such a quadric Q is expressible in the form

$$X^{n+1} = \tfrac{1}{2}H_{\sigma\tau}X^\sigma X^\tau + a_\rho X^\rho X^{n+1} + a(X^{n+1})^2, \tag{20}$$

the quadric of Čech given by (19) evidently belongs to this family. It is easily seen that the tangents to the intersection of V_n and Q drawn from O constitute a hypercone Γ^3_{n-1} of the third order

$$F_3 \equiv (\tfrac{1}{3}K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau})X^\sigma X^\tau X^\rho = 0. \tag{21}$$

The latter is apolar to the asymptotic hypercone

$$H_{\sigma\tau}X^\sigma X^\tau = 0 \tag{22}$$

if and only if

$$H^{\sigma\tau}(2K_{\sigma\tau i} - a_i H_{\sigma\tau} - 2a_\tau H_{\sigma i}) = 0 \qquad (i = i, \cdots, n),$$

namely,

$$a_i = \frac{2}{n+2}H^{\sigma\tau}K_{\sigma\tau i} \qquad (i = 1, \cdots, n). \tag{23}$$

In consequence, the quadric Q in consideration must belong to the pencil

$$X^{n+1} = \tfrac{1}{2}H_{\sigma\tau}X^\sigma X^\tau + \frac{2}{n+2}H^{\rho u}K_{\rho u\sigma}X^\sigma X^{n+1} + a(X^{n+1})^2, \tag{24}$$

called the Darboux pencil.[7] The equation (19) shows that the quadric of Čech is contained in the Darboux pencil.[8]

We now proceed to find the section V_ν of V_n produced by a space $[\nu + 1]$ of $\nu + 1$ dimensions through a given tangent space $[\nu]$ of ν dimensions at O. First it is convenient to take the given space $[\nu]$ through O for the space

$$X^{\nu+1} = 0, \cdots, X^{n+1} = 0 \qquad (\nu \geqq 1), \tag{25}$$

so that the $[\nu + 1]$ is then given by the equations

$$X^k = \lambda^k X^{n+1} \qquad (k = \nu + 1, \cdots, n). \tag{26}$$

In virtue of (1) and (26) there is no difficulty in showing that the section of V_n produced by (26) is a hypersurface V_ν in this $[\nu + 1]$, the expansion of V_ν at O being

$$\begin{aligned} X^{n+1} = \frac{1}{2}\sum_1^\nu H_{\sigma\tau}X^\sigma X^\tau &+ \frac{1}{3}\sum_1^\nu \bar{K}_{\sigma\tau\rho}X^\sigma X^\tau X^\rho \\ &+ \frac{1}{12}\sum_1^\nu \bar{H}_{\sigma\tau\rho u}X^\sigma X^\tau X^\rho X^u + \cdots, \end{aligned} \tag{27}$$

[7] J. Kanitani, *Géométrie différentielle projective des hypersurfaces*, Ryojun (1931), p. 33.
[8] Cf. *G. P. D.*, p. 617.

where the coefficients symmetrical in the subscripts are given by the following equations:

$$
(28)\quad \left\{\begin{aligned}
\bar K_{\sigma\tau\rho} &= K_{\sigma\tau\rho} + \tfrac12 H_{\tau\rho}\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^\alpha \\
&\qquad + \tfrac12 H_{\rho\sigma}\sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^\alpha + \tfrac12 H_{\sigma\tau}\sum_{\alpha=\nu+1}^{n} H_{\rho\alpha}\lambda^\alpha, \\
\bar H_{\sigma\tau\rho u} &= H_{\sigma\tau\rho u} + K_{\tau\rho u}\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^\alpha + K_{\rho u\sigma}\sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^\alpha + K_{u\sigma\tau}\sum_{\alpha=\nu+1}^{n} H_{\rho\alpha}\lambda^\alpha \\
&\quad + K_{\sigma\tau\rho}\sum_{\alpha=\nu+1}^{n} H_{u\alpha}\lambda^\alpha + \sum_{\alpha,\beta=\nu+1}^{n} (H_{\sigma\tau}H_{\rho\alpha}H_{u\beta} + H_{\sigma\rho}H_{u\alpha}H_{\tau\beta} \\
&\quad + H_{\sigma u}H_{\tau\alpha}H_{\rho\beta} + H_{\tau\rho}H_{u\alpha}H_{\sigma\beta} + H_{\tau u}H_{\sigma\alpha}H_{\rho\beta} + H_{\rho u}H_{\sigma\alpha}H_{\tau\beta})\lambda^\alpha\lambda^\beta \\
&\quad + \sum_{\alpha=\nu+1}^{n} (K_{\sigma\tau\alpha}H_{\rho u} + K_{\sigma\rho\alpha}H_{u\tau} + K_{\sigma u\alpha}H_{\tau\rho} + K_{\tau\rho\alpha}H_{\sigma u} + K_{\tau u\alpha}H_{\sigma\rho} \\
&\quad + K_{\rho u\alpha}H_{\sigma\tau})\lambda^\alpha + \tfrac12(H_{\sigma\tau}H_{\rho u} + H_{\sigma\rho}H_{\tau u} + H_{\rho\tau}H_{\sigma u}) \\
&\qquad \cdot \sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta}\lambda^\alpha\lambda^\beta \qquad (\sigma, \tau, \rho, u = 1, \cdots, \nu).
\end{aligned}\right.
$$

In order to find the quadric of Čech of V_ν at O we have only to replace in (19) n by ν; K_{ijk}, H_{ijkl}, etc. by $\bar K_{ijk}$, $\bar H_{ijkl}$, etc., remembering that in this case the summation must be of the terms obtained by giving the index each of its ν values $1, 2, \cdots, \nu$. Thus we have

$$
(29)\quad \begin{aligned}
&X^{n+1} - \frac12\sum_{\sigma,\tau=1}^{\nu} H_{\sigma\tau}X^\sigma X^\tau - \frac{2}{\nu+2}\sum_{\rho,\sigma,u=1}^{\nu} \bar H^{\rho u}\bar K_{\rho u\sigma}X^\sigma X^{n+1} \\
&\quad + \left\{\frac{2(\nu+3)}{\nu(\nu+2)^2}\sum_1^\nu \bar H^{\sigma\tau}\bar H^{\rho u}\bar K_{\sigma\tau r}\bar K_{\rho u s}\bar H^{rs} - \frac{1}{\nu(\nu+2)}\sum_1^\nu \bar H^{rs}\bar L_{rs}\right\}(X^{n+1})^2 = 0,
\end{aligned}
$$

where

$$
(30)\quad \begin{aligned}
\bar L_{rs} = \sum_1^\nu \bar H^{\sigma\tau}\bar H_{\sigma\tau rs} + \frac12\sum_1^\nu \bar H^{\sigma u,\tau\rho}(&\bar K_{\sigma\tau r}\bar K_{\rho u s} - \bar K_{u\tau r}\bar K_{\sigma\rho s} \\
&+ \bar K_{\sigma\tau s}\bar K_{\rho u r} - \bar K_{u\tau s}\bar K_{\sigma\rho r}).
\end{aligned}
$$

From (28) it follows that

$$
\sum_1^\nu \bar H^{\rho u}K_{\rho u\sigma} = \sum_1^\nu \bar H^{\rho u}K_{\rho u\sigma} + \tfrac12(\nu+2)\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^\alpha \qquad (\sigma = 1, \cdots, \nu).
$$

Hence, putting

$$
(31)\quad \left\{\begin{aligned}
\mathfrak{A} &= \sum_1^\nu \bar H^{\sigma\tau}\bar H^{\rho u}\bar H^{rs}\bar K_{\sigma\tau r}\bar K_{\rho u s}, \\
\mathfrak{B} &= \sum_1^\nu \bar H^{\sigma\tau}\bar H^{rs}\bar H_{\sigma\tau rs}, \\
\mathfrak{C} &= \sum_1^\nu \bar H^{\sigma u,\tau\rho}\bar H^{rs}(\bar K_{\sigma\tau r}\bar K_{\rho u s} - \bar K_{u\tau r}\bar K_{\rho\sigma s}),
\end{aligned}\right.
$$

the equation of the quadric of Čech in consideration can be written in the form

$$X^{n+1} - \frac{1}{2}\sum_1^\nu H_{\sigma\tau}X^\sigma X^\tau - \left\{\frac{2}{\nu+2}\sum_1^\nu \bar{H}^{\rho u}K_{\rho u\sigma}X^\sigma + \sum_{\sigma=1}^{\nu}\sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^\alpha X^\sigma\right\}X^{n+1} + \left\{\frac{2(\nu+3)}{\nu(\nu+2)^2}\mathfrak{A} - \frac{1}{\nu(\nu+2)}(\mathfrak{B}+\mathfrak{C})\right\}(X^{n+1})^2 = 0. \tag{32}$$

Since $\bar{K}_{\sigma\tau\rho}$ and $\bar{H}_{\sigma\tau\rho u}$ are respectively linear and quadratic in λ's and $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ given by (31) are all quadratic in these parameters, by eliminating them from (26) and (32) we may easily conclude that when the space $[\nu+1]$ turns about the fixed space $[\nu]$ given by (25) the quadric of Čech (32) describes a hyperquadric, which completes the proof of the theorem of Čech.

For the subsequent development it is, however, desirable to derive the explicit form for $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ and therefore the equation of the hyperquadric thus obtained, namely, the Moutard-Čech hyperquadric belonging to the given space $[\nu]$. The first two of them may easily be calculated by means of (28). Thus we obtain

$$\begin{aligned}\mathfrak{A} = {} & \sum_1^\nu \bar{H}^{\sigma\tau}\bar{H}^{\rho u}\bar{H}^{rs}K_{\sigma\tau r}K_{\rho us} \\ & + (\nu+2)\sum_1^\nu \bar{H}^{rs}\bar{H}^{\sigma\tau}K_{\sigma\tau r}\sum_{\alpha=\nu+1}^{n} H_{s\alpha}\lambda^\alpha \\ & + \frac{1}{4}(\nu+2)^2\sum_1^\nu \bar{H}^{rs}\sum_{\alpha,\beta=\nu+1}^{n} H_{r\alpha}H_{s\beta}\lambda^\alpha\lambda^\beta,\end{aligned} \tag{33}$$

$$\begin{aligned}\mathfrak{B} = {} & \sum_1^\nu \bar{H}^{\sigma\tau}\bar{H}^{\rho u}H_{\sigma\tau\rho u} + 4\sum_1^\nu \bar{H}^{\sigma\tau}\bar{H}^{\rho u}K_{\rho u\sigma}\sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^\alpha \\ & + 2(\nu+2)\sum_1^\nu \bar{H}^{rs}\sum_{\alpha,\beta=\nu+1}^{n} H_{r\alpha}H_{s\beta}\lambda^\alpha\lambda^\beta + 2(\nu+2)\sum_1^\nu \bar{H}^{\sigma\tau}\sum_{\alpha=\nu+1}^{n} K_{\sigma\tau\alpha}\lambda^\alpha \\ & + \frac{1}{2}\nu(\nu+2)\sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta}\lambda^\alpha\lambda^\beta,\end{aligned} \tag{34}$$

remembering that $H_{\sigma\tau}$, $K_{\sigma\tau\rho}$, $H_{\sigma\tau\rho u}$ are symmetrical in the subscripts and that

$$\sum_1^\nu \bar{H}^{\sigma\tau}H_{\sigma\rho} = \delta_\rho^\tau. \tag{35}$$

It remains for us to compute $\mathfrak{C}$. For this purpose we find it convenient to remark that

$$\left\{\begin{aligned} & \bar{H}^{\sigma u,\tau\rho} = -\bar{H}^{u\sigma,\tau\rho} = -\bar{H}^{\sigma u,\rho\tau} \\ & \bar{H}^{\sigma u,\tau\rho}H_{\tau r} = \bar{H}^{u\rho}\delta_r^\sigma - \bar{H}^{\sigma\rho}\delta_r^u, \\ & \bar{H}^{\sigma u,\tau\rho}H_{\sigma r} = \bar{H}^{u\rho}\delta_r^\tau - \bar{H}^{u\tau}\delta_r^\rho, \\ & \sum_1^\nu \bar{H}^{\sigma u,\tau\rho}(H_{\sigma\tau}H_{u\rho} - H_{u\tau}H_{\sigma\rho}) = 2\nu(\nu-1), \\ & \sum_1^\nu \bar{H}^{\sigma u,\tau\rho}(H_{\sigma r}H_{us} - H_{u r}H_{\sigma s}) = 2(\delta_r^\tau\delta_s^\rho - \delta_s^\tau\delta_r^\rho), \quad (r, s, \rho, \tau = 1, \cdots, \nu).\end{aligned}\right. \tag{36}$$

The first two relations are consequences of the definition of $\bar{H}^{\sigma u, \tau\rho}$, the next two follow from the expansions of the determinant involved in $\bar{H}^{\rho u}$, and the last two sets follow from the repeated utilization of Laplace expansions of the determinant $\bar{H}$ or from the preceding two equations. Some calculations suffice then to demonstrate that

$$
\begin{aligned}
\mathfrak{C} = \sum_1^{\nu} \bar{H}^{rs} \bar{H}^{\sigma u, \tau\rho}(K_{\sigma\tau r} K_{u\rho s} - K_{u\tau r} K_{\sigma\rho s}) \\
+ 2(\nu - 1) \sum_1^{\nu} \bar{H}^{rs} \bar{H}^{\sigma\tau} K_{\sigma\tau r} \sum_{\alpha=\nu+1}^{n} H_{s\alpha} \lambda^{\alpha} \\
+ \frac{1}{2} (\nu + 2)(\nu - 1) \sum_1^{\nu} \bar{H}^{rs} \sum_{\alpha,\beta=\nu+1}^{n} H_{r\alpha} H_{s\beta} \lambda^{\alpha} \lambda^{\beta}.
\end{aligned} \tag{37}
$$

Substituting (33), (34) and (37) into (32) and reducing, we arrive at the equation to the quadric of Čech of V_ν at O, namely,

$$
\begin{aligned}
X^{n+1} - \frac{1}{2} \sum_1^{\nu} H_{\sigma\tau} X^{\sigma} X^{\tau} - \left\{ \frac{2}{\nu + 2} \sum_1^{\nu} \bar{H}^{\rho u} K_{\rho u \sigma} X^{\sigma} + \sum_{\sigma=1}^{\nu} \sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha} \lambda^{\alpha} X^{\sigma} \right\} X^{n+1} \\
+ \left\{ \frac{2(\nu + 3)}{\nu(\nu + 2)^2} \sum_1^{\nu} \bar{H}^{rs} \bar{H}^{\rho u} \bar{H}^{\sigma\tau} K_{\sigma\tau r} K_{\rho u s} - \frac{1}{\nu(\nu + 2)} \sum_1^{\nu} \bar{H}^{\sigma\tau} \bar{L}_{\sigma\tau} \right. \\
+ \frac{4}{\nu(\nu + 2)} \sum_1^{\nu} \bar{H}^{\sigma\tau} \bar{H}^{\rho u} K_{\rho u \sigma} \sum_{\alpha=\nu+1}^{n} H_{\tau\alpha} \lambda^{\alpha} - \frac{2}{\nu} \sum_1^{\nu} \bar{H}^{\sigma\tau} \sum_{\alpha=\nu+1}^{n} K_{\sigma\tau\alpha} \lambda^{\alpha} \\
\left. - \frac{1}{2} \sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta} \lambda^{\alpha} \lambda^{\beta} \right\} (X^{n+1})^2 = 0
\end{aligned} \tag{38}
$$

where $\bar{L}_{\sigma\tau}(\sigma, \tau = 1, \cdots, \nu)$ are given by (4). Elimination of $\lambda^{\nu+1}, \cdots, \lambda^{n}$ between (26) and (38) immediately gives *the equation of the Moutard-Čech hyperquadric belonging to the space* (25):

$$
\begin{aligned}
X^{n+1} = \frac{1}{2} \sum_{\sigma,\tau=1}^{n} H_{\sigma\tau} \cdot X^{\sigma} X^{\tau} \\
+ \sum_{\rho=1}^{n} \left\{ \frac{2}{\nu} \sum \bar{H}^{\sigma\tau} K_{\sigma\tau\rho} - \frac{4}{\nu(\nu + 2)} \sum_{\sigma,\tau,u,v=1}^{\nu} H_{\sigma\rho} \bar{H}^{\sigma\tau} \bar{H}^{uv} K_{uv\tau} \right\} X^{\rho} X^{n+1} \\
- \left\{ \frac{2(\nu + 3)}{\nu(\nu + 2)} \sum_1^{\nu} \bar{H}^{\sigma\tau} \bar{H}^{\rho u} \bar{H}^{rs} K_{\sigma\tau r} K_{\rho u s} - \frac{1}{\nu(\nu + 2)} \sum_1^{\nu} \bar{H}^{\sigma\tau} \bar{L}_{\sigma\tau} \right\} (X^{n+1})^2.
\end{aligned} \tag{39}
$$

Several interesting theorems can be easily established. For example, take any point $P(\xi^1, \cdots, \xi^{\nu}, 0, \cdots, 0)$ in the given space $[\nu]$; the polar hyperplane (in $[\nu]$) of P with respect to the corresponding Moutard-Čech hyperquadric is evidently given by the equation

$$
X^{n+1} = \sum_{\sigma,\tau=1}^{\nu} H_{\sigma\tau} \xi^{\sigma} X^{\tau} + \frac{2}{\nu + 2} \sum_1^{\nu} \bar{H}^{\sigma\tau} K_{\sigma\tau\rho} \xi^{\rho} X^{n+1}. \tag{40}
$$

This hyperplane can, however, be obtained by the following method: Consider the quadric Q given by (20); the cubic hypercone Γ^3_{n-1} of equation (21) contains the space $[\nu]$ when and only when

$$\sum_1^\nu (\tfrac{1}{3}K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau})\xi^\sigma\xi^\tau\xi^\rho = 0,$$

that is,

$$K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau} - \tfrac{1}{2}a_\sigma H_{\tau\rho} - \tfrac{1}{2}a_\tau H_{\rho\sigma} = 0, \qquad (\sigma, \tau, \rho = 1, \cdots, \nu). \tag{41}$$

Multiplying (41) by $\bar{H}^{\sigma\tau}$ and summing for $\sigma, \tau = 1, \cdots, \nu$, we obtain

$$a_\rho = \frac{2}{\nu + 2}\sum_{\sigma,\tau=1}^{\nu} \bar{H}^{\sigma\tau}K_{\sigma\tau\rho}, \qquad (\rho = 1, \cdots, \nu). \tag{42}$$

The hyperquadric Q in this case becomes

$$\begin{aligned} X^{n+1} = \frac{1}{2}\sum_1^\nu H_{\sigma\tau}X^\sigma X^\tau + \frac{2}{\nu+2}\sum_1^\nu \bar{H}^{\sigma\tau}K_{\sigma\tau\rho}X^\rho X^{n+1} \\ + \sum_{\alpha=\nu+1}^{n} a_\alpha X^\alpha X^{n+1} + a(X^{n+1})^2, \end{aligned} \tag{43}$$

$a_{\nu+1}, \cdots, a_n$, a being arbitrary constants. It is obvious that the polar hyperplane of the point $P(\xi^1, \cdots, \xi^\nu, 0, \cdots, 0)$ is precisely (40). Hence we have the

THEOREM. *Suppose that a space $[\nu]$ through O be contained in the tangent hyperplane of a hypersurface V_n at O and that a hyperuqadric Q has at O a contact of the second order with V_n. If the cubic hypercone Γ^3_{n-1} formed by the tangents drawn from O to the intersection of V_n and Q contains the space $[\nu]$, then the polar hyperplane of any point in this $[\nu]$ with respect to Q must coincide with that of the same point with respect to the Moutard-Čech hyperquadric belonging to $[\nu]$.*

In particular when $\nu = 1$ we obtain a theorem due to Fubini.[9]

Let us now consider the hyperquadric (20) such that the corresponding Γ^3_{n-1} should pass through the space $[\nu]$ doubly. We can show that such a hyperquadric must belong to a pencil. In fact, for the hyperquadric Q in question the coefficients $a_\rho (\rho = 1, \cdots, \nu)$ are, as before, given by (42), and

$$\frac{\partial F_3}{\partial X^\rho} \equiv (K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau} - a_\sigma H_{\tau\rho})X^\sigma X^\tau = 0, \quad (\rho = 1, \cdots, n),$$

must be satisfied identically for any $\xi^1, \cdots, \xi^\nu$, and $\xi^\alpha = 0$, $(\alpha = \nu + 1, \cdots, n)$. Therefore

$$\sum_{\sigma,\tau=1}^{\nu} (K_{\sigma\tau\rho} - \tfrac{1}{2}a_\rho H_{\sigma\tau} - a_\sigma H_{\tau\rho})X^\sigma X^\tau \equiv 0, \quad (\rho = 1, \cdots, n),$$

or

$$2K_{\sigma\tau\rho} - a_\rho H_{\sigma\rho} - a_\sigma H_{\tau\rho} - a_\tau H_{\sigma\rho} = 0 \quad (\sigma, \tau = 1, \cdots, \nu; \rho = 1, \cdots, n), \tag{44}$$

[9] Cf. *G. P. D.*, pp. 617–618.

whence

$$a_\rho = \frac{2}{\nu}\sum_{\sigma,\tau=1}^{\nu} \bar{H}^{\sigma\tau} K_{\sigma\tau\rho} - \frac{4}{\nu(\nu+2)} \sum_{\sigma,\tau,u,v=1}^{\nu} H_{\rho\sigma}\bar{H}^{\sigma\tau}\bar{H}^{uv}K_{uv\tau} \quad (\rho = 1, \cdots, n). \tag{45}$$

Hence we obtain a pencil of hyperquadrics

$$\begin{aligned} X^{n+1} &= \frac{1}{2}\sum_{\sigma,\tau=1}^{n} H_{\sigma\tau}X^\sigma X^\tau \\ &\quad + \sum_{\rho=1}^{n}\left\{\frac{2}{\nu}\sum_{\sigma,\tau=1}^{\nu}\bar{H}^{\sigma\tau}K_{\sigma\tau\rho} - \frac{4}{\nu(\nu+2)}\sum_{\sigma,\tau,u,v=1}^{\nu} H_{\rho\sigma}\bar{H}^{\sigma\tau}\bar{H}^{uv}K_{uv\tau}\right\} X^\rho X^{n+1} \\ &\quad + a(X^{n+1})^2, \end{aligned} \tag{46}$$

a being a parameter. We shall call (46) *the pencil of Moutard-Čech* belonging to the given space $[\nu]$, because each of these pencils contains a Moutard-Čech hyperquadric (39). Thus we have proved the following

THEOREM. *If the cubic hypercone Γ^3_{n-1} passes through the space $[\nu]$ doubly, then the hyperquadric Q must necessarily be in a pencil of Moutard-Čech belonging to the space $[\nu]$.*

That the converse of this theorem is not necessarily true may be seen by considering a V_n in which $H_{\sigma\tau} = \delta_{\sigma\tau}$ and $K_{\rho\sigma\tau} \neq 0$ for ρ, σ, τ all different. Then even if a_ρ is defined by (45), the right member of (44) becomes $K_{\rho\sigma\tau} \neq 0$ for ρ, σ, τ all different, and thus (44) cannot hold.

This generalizes a theorem for the case $\nu = 1$.[10]

The hyperquadric of the pencil (46) possesses another definition, as we will show below.

From (27) it is easily seen that all of sections V_ν produced by spaces $[\nu + 1]$ through the space $[\nu]$ have the asymptotic hypercone at O in common, namely,

$$\sum_{\sigma,\tau=1}^{\nu} H_{\sigma\tau}X^\sigma X^\tau = 0, \qquad X^{\nu+1} = 0, \cdots, X^{n+1} = 0, \tag{47}$$

which may be obtained as the intersection of the space $[\nu]$ with the asymptotic hypercone (22).

In the tangent hyperplane $X^{n+1} = 0$ of V_n at O a space $[n - \nu - 1]$ is taken such that it is skew with the space $[\nu]$. The projection of the hypercone (47) from this $[n - \nu - 1]$ is *apolar* to the cubic hypercone Γ^3_{n-1} if and only if

$$\sum_{\sigma,\tau=1}^{\nu} \bar{H}^{\sigma\tau}(2K_{\sigma\tau\rho} - a_\rho H_{\sigma\tau} - a_\sigma H_{\rho\tau} - a_\tau H_{\rho\sigma}) = 0, \qquad (\rho = 1, \cdots, n),$$

whence we obtain (45) and consequently the pencil of hyperquadrics (46). Hence we have the

THEOREM. *If the cubic hypercone Γ^3_{n-1} be apolar to the quadratic cone obtained by projecting the common asymptotic hypercone of various V_ν's from a space*

[10] Cf. my paper: *Plane sections through an ordinary point of a hypersurface*, to be published in Revista, Tucumán.

$[n - \nu - 1]$ *contained in the tangent hyperplane of* V_n *at* O, *but skew with the given* $[\nu]$, *then the corresponding hyperquadric* Q *must belong to the pencil of Moutard-Čech, and conversely.*

We have shown the truth of this theorem in the case $\nu = 2$.[11]

Among various sections of a hyperquadric Q of the equation (20) produced by spaces (26) of $\nu + 1$ dimensions through the given space $[\nu]$ it may happen that the section just coincides with the quadric of Čech of the section V_ν. In virtue of (20) and (26) it is easily shown that the section of Q is

$$\begin{aligned} 2X^{n+1} = \sum_{\sigma,\tau=1}^{\nu} H_{\sigma\tau} X^\sigma X^\tau + 2\sum_{\sigma=1}^{\nu}\left\{a_\sigma + \sum_{\alpha=\nu+1}^{n} H_{\sigma\alpha}\lambda^\alpha\right\} X^\sigma X^{n+1} \\ + 2\left(a + \sum_{\alpha=\nu+1}^{n} a_\alpha \lambda^\alpha\right)(X^{n+1})^2. \end{aligned} \tag{48}$$

In order that the latter should represent the quadric of Čech (38), the necessary and sufficient conditions are

$$a_\sigma = \frac{2}{\nu+2}\sum_{\rho,u=1}^{\nu} \bar{H}^{\rho u} K_{\rho u \sigma}, \qquad (\sigma = 1, \cdots, \nu), \tag{49}$$

$$\begin{aligned} a + \sum_{\alpha=\nu+1}^{n} a_\alpha \lambda^\alpha = \frac{1}{2}\sum_{\alpha,\beta=\nu+1}^{n} H_{\alpha\beta}\lambda^\alpha\lambda^\beta + \frac{2}{\nu}\sum_{1}^{\nu} \bar{H}^{\sigma\tau} \sum_{\alpha=\nu+1}^{n} K_{\sigma\tau\alpha}\lambda^\alpha \\ + \frac{1}{\nu(\nu+2)}\sum^{\nu} \bar{H}^{\sigma\tau}\bar{L}_{\sigma\tau} \\ - \frac{4}{\nu(\nu+2)}\sum_{1}^{\nu} \bar{H}^{\sigma\tau}\bar{H}^{\rho u} K_{\rho u \sigma}\sum_{\alpha=\nu+1}^{n} H_{\tau\alpha}\lambda^\alpha \\ - \frac{2(\nu+3)}{\nu(\nu+2)^2}\sum_{1}^{\nu} \bar{H}^{rs}\bar{H}^{\sigma\tau}\bar{H}^{\rho u} K_{\sigma\tau r} K_{\rho u s}. \end{aligned} \tag{50}$$

I have not been able to find the *geometrical meaning* of (49) and (50). We remark that if the hypercone Γ^3_{n-1} passes through $[\nu]$ then (42), which is identical with (49), holds. However, since (49), or (42), does not imply (41)—as can be seen from the counterexample following the next to the last theorem above—(49) does not imply that Γ^3_{n-1} passes through $[\nu]$. We also remark that in the case $\nu = 1$ the equation (50) shows that *the plane* (26) *should osculate a curve contained in the intersection of* V_n *and* Q.[12]

It is of some interest to investigate all the quadrics of Čech of sections V_ν which lie on a fixed hyperquadric Q having at O a contact of the second order with V_n. Observing that a space $[\nu]$ through O and in the tangent hyperplan of V_n at O contains in general $\nu(n - \nu)$ parameters and that the conditions (49) and (50) are $\nu + 1$ in number, we may conclude that *on* Q *there are* $\infty^{(\nu+1)(n-\nu-1)}$ *quadrics of Čech of* ν *dimensions.*

[11] Cf. my paper, loc. cit.[10].

[12] Cf. my paper, loc. cit.[10].

For the sake of completeness we shall find the equation of the Moutard-Čech hyperquadric belonging to a general space $[\nu]$ through O and in the tangent hyperplane of V_n at O.

For this purpose it is convenient to express the equations to the space $[\nu]$ in the form

$$X^\alpha = \sum_{\sigma=1}^{\nu} d_\sigma^\alpha X^\sigma \qquad (\alpha = \nu + 1, \cdots, n), \qquad X^{n+1} = 0 \tag{51}$$

or

$$\bar{X}^k = 0 \qquad (k = \nu + 1, \cdots, n), \qquad X^{n+1} = 0 \tag{52}$$

if we put

$$\bar{X}^k = X^k - \sum_{\sigma=1}^{\nu} d_\sigma^k X^\sigma \qquad (k = \nu + 1, \cdots, n). \tag{53}$$

Introduce

$$\bar{X}^j = X^j \qquad (j = 1, \cdots, \nu), \qquad \bar{X}^{m+1} = X^{n+1}, \tag{54}$$

so that

$$X^k = \bar{X}^k + \sum_{\sigma=1}^{\nu} d_\sigma^k X^\sigma \qquad (k = \nu + 1, \cdots, n). \tag{55}$$

Substituting (54), (55) into (1) and rearranging, we obtain

$$\begin{aligned}\bar{X}^{n+1} &= \frac{1}{2}\sum_1^n \bar{H}_{\sigma\tau}\bar{X}^\sigma\bar{X}^\tau + \frac{1}{3}\sum_1^n \bar{K}_{\sigma\tau\rho}\bar{X}^\sigma\bar{X}^\tau\bar{X}^\rho \\ &\qquad + \frac{1}{12}\sum \bar{H}_{\sigma\tau\rho u}\bar{X}^\sigma\bar{X}^\tau\bar{X}^\rho\bar{X}^u + \cdots,\end{aligned} \tag{56}$$

where

$$\left\{\begin{aligned} \bar{H}_{\sigma\tau} &= H_{\sigma\tau} + \sum_{\alpha=\nu+1}^n H_{\sigma\alpha} d_\tau^\alpha + \sum_{\alpha=\nu+1}^n H_{\tau\alpha} d_\sigma^\alpha + \sum_{\alpha,\beta=\nu+1}^n H_{\alpha\beta} d_\sigma^\alpha d_\tau^\beta, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\sigma, \tau = 1, \cdots, \nu), \\ \bar{H}_{\sigma\alpha} &= H_{\sigma\alpha} + \sum_{\beta=\nu+1}^n H_{\alpha\beta} d_\sigma^\beta, \qquad (\sigma = 1, \cdots, \nu;\ \alpha = \nu + 1, \cdots, n), \\ \bar{H}_{\alpha\beta} &= H_{\alpha\beta}, \qquad\qquad\qquad\qquad (\alpha, \beta = \nu + 1, \cdots, n);\end{aligned}\right. \tag{57}$$

and similarly one can obtain explicit expressions for $\bar{K}_{\rho\sigma\tau}$ and $\bar{H}_{\rho\sigma\tau u}$ in terms of $K_{\rho\sigma\tau}$, $H_{\rho\sigma\tau u}$, and d_ρ^α.

By virtue of (39) the equation of the Moutard-Čech hyperquadric belonging to (51) is of the form

$$\begin{aligned}\bar{X}^{n+1} &= \frac{1}{2}\sum_1^n \bar{\bar{H}}_{\sigma\tau}\bar{X}^\sigma\bar{X}^\tau \\ &\quad + \sum_{\rho=1}^n \left\{\frac{2}{\nu}\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{K}_{\sigma\tau\rho} - \frac{4}{\nu(\nu+2)}\sum_1^\nu \bar{\bar{H}}_{\sigma\rho}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{uv}\bar{K}_{uv\tau}\right\}\bar{X}^\rho\bar{X}^{n+1} \\ &\quad - \left\{\frac{2(\nu+3)}{\nu(\nu+2)^2}\sum_1^\nu \bar{\bar{H}}^{rs}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{\rho u}\bar{K}_{\sigma\tau r}\bar{K}_{\rho us} - \frac{1}{\nu(\nu+2)}\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{L}_{\sigma\tau}\right\}(\bar{X}^{n+1})^2\end{aligned} \tag{58}$$

where $\bar{\bar{H}}^{\sigma\tau}$ is related to $\bar{H}_{\sigma\tau}$ as $\bar{H}_{\sigma\tau}$ was to $H_{\sigma\tau}$. On account of (53), (54), (57) the equation (58) may be written as

$$
\begin{aligned}
X^{n+1} = {} & \frac{1}{2}\sum_1^n H_{\sigma\tau}X^\sigma X^\tau + \frac{2}{\nu}\sum_{\rho=1}^n\Big\{\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{K}_{\sigma\tau\rho} - \frac{2}{\nu+2}\sum_1^\nu \bar{H}_{\sigma\rho}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{uv}\bar{K}_{uv\tau} \\
& - \sum_{\alpha=\nu+1}^n\Big(\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{K}_{\sigma\tau\alpha} - \frac{2}{\nu+2}\sum_1^\nu \bar{H}_{\sigma\alpha}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{uv}\bar{K}_{uv\tau}\Big)d_\rho^\alpha\Big\}X^\rho X^{n+1} \\
& + \frac{1}{\nu(\nu+2)}\Big\{\sum_1^\nu \bar{\bar{H}}^{\sigma\tau}\bar{L}_{\sigma\tau} - 2(\nu+3)\sum_1^\nu \bar{\bar{H}}^{rs}\bar{\bar{H}}^{\sigma\tau}\bar{\bar{H}}^{\rho u}\bar{K}_{\sigma\tau r}\bar{K}_{\rho us}\Big\}(X^{n+1})^2.
\end{aligned}
\tag{59}
$$

Hence we have the

THEOREM. *The Moutard-Čech hyperquadrics belonging to two different spaces through O and in the tangent hyperplane of V_n at O intersect in the asymptotic hypercone of V_n at O and a hyperquadric of $n - 1$ dimensions.*

We shall consider now the intersection of the quadric of Čech of V_n with the Moutard-Čech hyperquadric (59) or what amounts to the same thing, (58). Since (19) remains unaltered when H, K, X are replaced by $\bar{H}$, $\bar{K}$, $\bar{X}$ respectively, the intersection in question consists of the asymptotic hypercone and a hyperquadric Q_{n-1}. The hyperplane on which Q_{n-1} lies passes through the space (52) if and only if

$$
\frac{1}{\nu+2}\sum_1^\nu \bar{\bar{H}}^{\rho u}\bar{K}_{\rho u\sigma} - \frac{1}{n+2}\sum_1^n \bar{H}^{\rho u}\bar{K}_{\rho u\sigma} = 0, \qquad (\sigma = 1, \cdots, \nu).
\tag{60}
$$

But the Darboux tangents of V_n at 0 constitute a cubic hypercone with the equations

$$
\sum_1^n \mathfrak{K}_{\rho u\sigma}\bar{X}^\rho\bar{X}^u\bar{X}^\sigma = 0, \qquad \bar{X}^{n+1} = 0,
\tag{61}
$$

where we have placed[13]

$$
\begin{aligned}
\mathfrak{K}_{\rho u\sigma} = \bar{K}_{\rho u\sigma} - \frac{1}{n+2}\Big(&\bar{H}_{u\sigma}\sum_1^n \bar{H}^{rs}\bar{K}_{rs\rho} \\
&+ \bar{H}_{\rho\sigma}\sum_1^n \bar{H}^{rs}\bar{K}_{rsu} + \bar{H}_{\rho u}\sum_1^n \bar{H}^{rs}\bar{K}_{rs\sigma}\Big).
\end{aligned}
\tag{62}
$$

In order that this hypercone should contain the space (52) it is necessary and sufficient that

$$
\mathfrak{K}_{\rho u\sigma} = 0 \qquad (\rho, u, \sigma = 1, \cdots, \nu)
\tag{63}
$$

Multiplying (63) by $\bar{H}^{\rho u}$ and summing up with regard to $\rho, u = 1, \cdots, \nu$, we are led to (60). Hence there follows the

[13] Cf. Kanitani, loc. cit., p. 34.

THEOREM. *At a generic point O of a hypersurface V_n the quadric of Čech (or any quadric in the Darboux pencil) and the Moutard-Čech hyperquadric (or any hyperquadric in the pencil of Moutard-Čech) belonging to a tangent space $[\nu]$ intersect in the asymptotic hypercone of V_n and another hyperquadric of $n - 1$ dimensions. The hyperplane through the latter hyperquadric contains the space $[\nu]$ when and only when the space $[\nu]$ belongs to the cubic hypercone of Darboux tangents at O.*

Thus is generalized the theorem of Bompiani[14] concerning Moutard quadrics of a surface in ordinary space.

NATIONAL UNIVERSITY OF CHEKIANG,
KWEICHOW, CHINA.

[14] E. Bompiani, *Contributi alla geometria proiettiva-differenziale di una superficie*, Bollettino della Unione Matematical Italiana, 3 (1924), p. 97.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

THE CLASS-RING IN MULTIPLICATIVE SYSTEMS

By A. R. Richardson

(Received October 15, 1942)

Notation. Σ denotes a system closed to a binary operation; $a, b, c, \cdots$, are elements of E-subsets $A, B, C, \cdots$ of Σ; $u, v, w, \cdots$, are elements of Σ. The term E-set implies that the classes $C_w = w \cdot A$ are disjoint, the relation between the elements in C_w being an equivalence relation R_A. Hence if w be any element in Σ, a_1, a_2 any elements in A then there exists an element a such that $w = w \cdot a$; $w = (w \cdot a_1) \cdot a_2$; $(w \cdot a_1) \cdot a_2 = w \cdot a$.

These classes will be regarded as the elements of a multiplicative system T_A in which multiplication will be denoted by $\times$. In group theory $\times$ may be derived from $\cdot$, e.g. $a \times b$ may denote any one of $a \cdot b$, $b \cdot a$, $a \cdot b \cdot a^{-1} \cdot b^{-1}$, $b^{-1} \cdot a \cdot b$ etc. On the other hand $\times$ may be defined without reference to $\cdot$ as in the theory of rings of systems closed to two operations, e.g. $ab - ba$, $ab + ba$.[1]

In order to establish an automorphism between the structure (lattice) of equivalence relations R_A, in which A ranges over all E-subsets of Σ, and those quotients of E-sets having the same R_A, we shall assume:

Cancellation Law I. *If for any element u we have $u \cdot a = u \cdot b$ then there exists a d in $D = A \cap B$ such that $u \cdot a = u \cdot b = u \cdot d$, i.e. we assume that D is non-void.*

Associative Laws. $w \cdot (a \cdot b) = (w \cdot a) \cdot \bar{b}$; $(w \cdot a) \cdot b = w \cdot (a \cdot \bar{b})$ *where $\bar{b}$ is in B. Hence $A \cdot (B \cdot C) = (A \cdot B) \cdot C$.*

The product $w \cdot a_i \times v \cdot a_j$ is in Σ and so falls into one and only one class $C_u = u \cdot A$. In ordinary algebra several classes arise, viz.:

(a) All elements $w \cdot A \times v \cdot A$ are in the same class $(w \times v) \cdot A$. This holds in the theory of normal *co*-set expansions in groups and semi-groups.

(b) All elements $w \cdot A \times v \cdot A$ are in the same class $u \cdot A$ where $u \neq w \times v$. This case cannot occur here for since A is an E-set, right units a_w, a_v exist such that $w = w \cdot a_w$ and $v = v \cdot a_v$. Hence $w \times v$ is in $w \cdot A \times v \cdot A$ which product must therefore be in $(w \times v) \cdot A$.

(c) The elements $w \cdot A \times v \cdot A$ fall into different classes, e.g. the classes of conjugate elements in group theory.

We shall denote unions and cross-cuts with respect to $\cdot$ and $\times$ by $A \overset{\cdot}{\cup} B$, $A \overset{\times}{\cup} B$, $A \overset{\cdot}{\cap} B$, $A \overset{\times}{\cap} B$ respectively.

1. $C_w \times C_v \leqq C_{w \times v}$.

This is equivalent to

$$(w \cdot A) \times (v \cdot A) \leqq (w \times v) \cdot A \tag{1}$$

[1] A. R. Richardson, *Congruences in Multiplicative Systems*. To appear in Proc. London Math. Soc.

i.e. given a_i, a_j any elements of A; w, v, any elements of Σ then $a \epsilon A$ exists such that

$$(w \cdot a_i) \times (v \cdot a_j) = (w \times v) \cdot a. \tag{2}$$

Since A is an E-set $a_w \epsilon A$ exists such that $w \cdot a_w = w$. Hence there are the following special instances of the distributive law (2):

$$w \times (v \cdot a_j) = (w \times v) \cdot a \tag{3}$$

$$(w \cdot a_i) \times v = (w \times v) \cdot a.^2 \tag{4}$$

If $(w \cdot A) \times (v \cdot A) = (w \times v) \cdot A$ there are other special instances:

$$w \times (a_i \cdot a_j) = (w \times a_i) \cdot a_j,$$

$$(w \times a_i) \cdot a_j = (w \cdot a_k) \times (a_i \cdot a_m) = (w \cdot a_k) \times a_n,$$

$$(w \cdot a_i) \times a_j = (w \cdot a_i) \times (a_j \cdot a_l) = (w \times a_j) \cdot a.$$

THEOREM 1. *If the classes mod R_A and mod R_B both satisfy condition* (2) *then so do those mod $R_{A \dot{\cup} B}$ and mod $R_{A \dot{\cap} B}$.*

Let $D = A \dot{\cap} B$, supposedly non-void, and let d_i, d_j be any elements in D. Then $(w \cdot d_i) \times (v \cdot d_j) = (w \cdot a_i) \times (v \cdot a_j) = (w \times v) \cdot a$. Similarly $(w \cdot d_i) \times (v \cdot d_j) = (w \times v) \cdot b$. Hence, by cancellation law I, there exists d in D such that $(w \cdot d_i) \times (v \cdot d_j) = (w \times v) \cdot d$, i.e. (2) holds for $D = A \dot{\cap} B$. Next, owing to the associative laws, the elements of $M = A \dot{\cup} B$ are of one of the forms a, b, $a \cdot b$, $b \cdot a$. Take these in turn:

$$(w \cdot a_i) \times (v \cdot b_j) = \{(w \cdot a_i) \times v\} \cdot b = \{(w \times v) \cdot a\} \cdot b = (w \times v) \cdot (a \cdot \bar{b})$$
$$= (w \times v) \cdot m.$$

$$\{w \cdot (a_1 \cdot b_1)\} \times \{v \cdot (a_2 \cdot b_2)\} = \{(w \cdot a_1) \cdot b_3\}\{(v \cdot a_2) \cdot b_4\} = \{(w \cdot a_1) \times (v \cdot a_2)\} \cdot b$$
$$= \{(w \times v) \cdot a\} \cdot b = (w \times v) \cdot (a \cdot b_5) = (w \times v) \cdot m.$$

$$\{w \cdot (a_1 \cdot b_1)\} \times (v \cdot (b_2 \cdot a_2)\} = \{w \cdot (a_1 \cdot b_1)\} \times \{(v \cdot b_2) \cdot a_3\}$$
$$= [\{w \cdot (a_1 \cdot b_1)\} \times (v \cdot b_2)] \cdot a = [\{(w \cdot a_1) \cdot b_3\} \times \{v \cdot b_2\}] \cdot a = [\{(w \cdot a_1) \times v\} \cdot b] \cdot a$$
$$= [\{(w \times v) \cdot a_4\} \cdot b] \cdot a = (w \times v) \cdot m.$$

Hence (2) holds for $M = A \dot{\cup} B$.

COROLLARY 1. *We have also proved that $A \cdot B$, defined as the set of all elements of Σ of the form $a \cdot b$, also satisfies* (2).

COROLLARY 2. *If A is a normal E-set of Σ with respect to $\cdot$, i.e. if $v \cdot A = A \cdot v$ for all v in Σ, then* (2) *is satisfied if $\times$ is taken to be $\cdot$.*

If (2) is satisfied it is customary to define $C_w \times C_v$ to be equal to $C_{w \times v}$, for every

[2] These laws resemble those of D. C. Murdock and O. Ore, *On generalized rings*. Am. Journ. of Math. vol. LXIII No. 1. 1941.

element in $C_w \times C_v$ is in $C_{w\times v}$. In such cases the correspondence $w \to w\cdot A$ between Σ and T_A is a homomorphism.

Although in many instances we require only that $(w\cdot A) \times (v\cdot A) \leqq (w \times v)\cdot A$ in others it is necessary that $(w\cdot A) \times (v\cdot A) = (w \times v)\cdot A$. We therefore examine the condition $(w \times v)\cdot A \leqq (w\cdot A) \times (v\cdot A)$. For this to hold it is necessary that given w, v, a there shall exist a_i and a_j such that

$$(w \times v)\cdot a = (w\cdot a_i) \times (v\cdot a_j). \tag{5}$$

THEOREM 2. *If the classes mod R_A and mod R_B satisfy* (5) *then so do those mod $R_{A\dot{\cup}B}$.*

The elements of $A \dot{\cup} B$ are of one of the forms a, b, $a\cdot b$, $b\cdot a$. Consider

$$\begin{aligned}(w \times v)\cdot m &= (w \times v)\cdot(a\cdot b) = [(w \times v)\cdot a]\cdot b_2 = [(w\cdot a_i) \times (v\cdot a_j)]\cdot b_2\\ &= \{(w\cdot a_i)\cdot b_i\} \times \{(v\cdot a_j)\cdot b_j\} = \{w\cdot(a_i\cdot \bar{b}_i)\} \times \{v\cdot(a_j\cdot\bar{b}_j)\}\\ &= (w\cdot m_i) \times (v\cdot m_j)\end{aligned}$$

i.e. (5) holds. Similarly it holds for the case $m = b\cdot a$ and therefore in all cases.

In order that (5) shall hold for $R_{A\dot{\cap}B}$ a new form of the cancellation law is necessary.

CANCELLATION LAW II. *If $(w\cdot a_i) \times (v\cdot a_j) = (w\cdot b_i) \times (v\cdot b_j)$ then there exist in $D = A \dot{\cap} B$ d_i and d_j such that the above products equal $(w\cdot d_i) \times (v\cdot d_j)$.*

THEOREM 3. *If the cancellation law II holds and if R_A, R_B satisfy condition* (5) *then so does $R_{A\dot{\cap}B}$.*

$(w \times v)\cdot d = (w\cdot a_i) \times (v\cdot a_j) = (w\cdot b_i) \times (v\cdot b_j)$ since d is in both A and B. Hence, from cancellation law II, d_i, d_j exist in D such that $(w \times v)\cdot d = (w\cdot d_i) \times (v\cdot d_j)$, i.e. (5) holds in $R_{A\dot{\cap}B}$.

THEOREM 4. *If Σ is homogeneous with respect to $\times$ and if cancellation law II and laws* (2) *and* (5) *hold for A and B then the cancellation law I also holds.*

Let z be any element in Σ. Since Σ is homogeneous, w and v exist such that $z = w \times v$. Let $z\cdot a = z\cdot b$ then a_i, a_j, b_i, b_j exist such that $z\cdot a = z\cdot b = (w\cdot a_i) \times (v\cdot a_j) = (w\cdot b_i) \times (v\cdot b_j)$ and by law II, d_i, d_j exist in $D = A \dot{\cap} B$ such that $z\cdot a = z\cdot b = (w\cdot d_i) \times (v\cdot d_j)$ and by law (2), there exists d in D such that these products are equal to $(w \times v)\cdot d = z\cdot d$, i.e. cancellation law I holds.

THEOREM 5. *If Σ is homogeneous with respect to Σ and if cancellation law I holds as well as laws* (2) *and* (5), *then the cancellation law II holds.*

Let $(w\cdot a_i) \times (v\cdot a_j) = (w\cdot b_i) \times (v\cdot b_j)$ then by (2) a and b exist such that $(w \times v)\cdot a = (w \times v)\cdot b$. Hence by cancellation law I, d exists in $D = A \dot{\cap} B$ such that $(w \times v)\cdot a = (w \times v)\cdot b = (w \times v)\cdot d$ and, since (5) also holds, this is equal to $(w\cdot d_i) \times (v\cdot d_j)$, i.e. cancellation law II holds.

In general the E-sets do not form a Dedekind Structure (modular lattice). If however the equation $a_1 = a_2\cdot x$ is always solvable in an E-set, then

THEOREM 6. *The E-sets form a Dedekind structure with respect to $\cdot$, i.e. if $A < C$ then $C \dot{\cap} (A \dot{\cup} B) = A \dot{\cup} (B \dot{\cap} C)$.*

By the associate laws the elements of $A \dot{\cup} B$ are of one of the forms a, b, $a \cdot b$, $b \cdot a$.

Let $c = b \cdot a$. Then $\bar{a}$ exists in A and therefore in C such that $c \cdot \bar{a} = b$. Hence b is in C and therefore in $B \dot{\cap} C$, i.e. $e \in A \dot{\cup} (B \dot{\cap} C)$.

Let $c = a \cdot b$ then $\bar{c}$ exists in C such that $c = a \cdot \bar{c}$, i.e. $c = a \cdot \bar{c} = a \cdot b$. Hence d exists in $B \dot{\cap} C$ such that $c = a \cdot d$, i.e. $c \in A \dot{\cup} (B \dot{\cap} C)$.

Corollary. *The mutually permutable E-sets A, B, $\cdots$ for which $a_1 = a_2 \cdot x$, $b_1 = b_2 \cdot x$ have solutions, form a Dedekind structure.*

It is desirable to express the conditions that every element in a product class occurs equally often in the product but, apart from the very restrictive conditions such as those in group theory, the necessary qualifications are complicated and are not inserted here. Instead we shall assume the $C_i C_j = \Sigma c_{ijk} C_k$.

2. The characteristic equations in the class-ring

Let $X = \sum x_i C_i$, $i = 1, 2, \cdots, n$, the x's being indeterminates in a ring K not of characteristic two. Then

$$X \times Y = \sum x_i C_i \times \sum y_j C_j = \sum_{i,j} x_i y_j C_i C_j = (x_i)(\sum_j y_j c_{kjm})(C_n)$$

where (x_i) is a one-rowed matrix and C_n a one-columned matrix and $\Delta_Y = (\sum_j y_j c_{kjm})$ is an n-matrix in which k denotes the rows and m the columns. Hence

$$(x_i)(C_p)(y_j)(C_q) = (x_i)\Delta_Y(C_n).$$

Similarly

$$\begin{aligned} X \times Y &= (C_i)(\sum_j x_j c_{jmk})(y_n) \\ &= (y_i)(\sum x_j c_{jkm})(C_n) \\ &= (C_i)(\sum y_j c_{mjk})(x_n). \end{aligned}$$

Evidently $(X \times Y) \times Z = (x_i)\Delta_Y \Delta_Z(C_n)$ and $X \times (Y \times Z) = (x_i)\Delta_{Y \times Z}(C_n)$.

Hence the condition for the class-ring to be associative is $\Delta_{Y \times Z} = \Delta_Y \Delta_Z$. If this holds then the correspondence $Y \to \Delta_Y$ between the class-ring and the matrices in a homomorphism m in which the class multiplication x corresponds to matrix multiplication and addition to addition.

Let $X^{(n+1)} = X^{(n)} \times X$. Then $X^{(n+1)} = (x_i)\Delta_X^n(C_n)$. Hence if Δ_X satisfies the characteristic equation $Z^n + p_1 Z^{n-1} + \cdots + p_n = 0$, then X satisfies a characteristic equation

$$X^{(n+1)} + p_1 X^{(n)} + \cdots + p_n X = 0.$$

Evidently the p_i are homogeneous and of degree i in the x's and the elements of Δ_X are linear in the x's and integral in the c_{ijk}.

In group theory and in the theory of matrix representations of algebras the

regular representation is important. From the present point of view this corresponds to Δ_E where R_E is the unit equivalence relation in which each class consists of one and only one element of Σ.

THEOREM 7. *If every element in a class-product occurs equally often in its class then the characteristic polynomial of the general number in the class-ring has a linear factor* $Z - \sum x_i\rho_i$ *where* ρ_i *is the number of elements of* Σ *in* C_i.

Since every element occurs equally often in its class,

$$\rho_i\rho_j = \sum c_{ijk}\rho_k .$$

Multiply the columns of Δ_Z by $\rho_1, \rho_2, \cdots \rho_n$ respectively and add to the first. Then the γ^{th} row has in the first column the element

$$\sum_{j,m} x_j C_{\gamma jm}\,\rho_m - Z\rho_\gamma = \sum_j z_j\,\rho_\gamma\,\rho_j - Z\rho_\gamma = \rho_\gamma\left[\sum z_j\,\rho_j - Z\right]. \tag{6}$$

Hence $Z - \sum z_j\rho_j$ is a factor of the characteristic polynomial Δ_Z and it is linear in the indeterminates z_j.

In group theory $a\cdot b = b^{-1}\cdot a\cdot b$ and $a \times b = a\cdot b$. Every element in a class occurs equally often in a class-product and the characteristic equation of the general number in the class-ring splits into linear factors

$$\prod_i \left[Z - \sum_j z_j\,\rho_j\,X_j^{(i)}/x_0^{(i)}\right]$$

where the $X_j^{(i)}$ are the group characteristics. The group character $(1, 1, \cdots, 1)$ arises from (6).

In the general case the characteristic polynomial does not split into linear factors although it is known to do so when the class-ring is commutative.

THEOREM 8. *If* Σ *is an E-set with respect to* $\times$ *then* $\Delta_Z = [Z - \sum z_i\rho_i]^n$ *is a complete* n^{th} *power.*

For, Σ being an E-set, the elements which appear in the multiplicative table of Σ as multiples of C_i, $i = 1, 2, \cdots, n$, are in C_i. Hence $c_{ijk} = 0$, $k \neq i$ and $C_{iji} = \rho_j$, $k = i$. Hence $\Delta_Z = [Z - \sum z_i\rho_i]^n$.

THEOREM 9. *If* $R_A \subset R_B$ *then* $|\,\Delta_B\,|$ *is a divisor of* $|\,\Delta_A'\,|$ *where* Δ_A' *is what* Δ_A *becomes when certain of the a's are equal.*

Since $R_A \subset R_B$, $B_i = \sum C_s$ and therefore $B_iB_j = \sum b_{ijk}C_k = \sum C_nC_m = \sum c_{nmk}C_k$ i.e. $\sum b_{ijk} = \sum C_p = \sum c_{nmk}C_k$. Let $B_1 = C_1 + \cdots + C_{s_1}$; $B_2 = C_{s_1+1} + \cdots + C_{s_2}, \cdots$
then

$$B_iB_j = \sum (C_{s_{i-1}+1} + \cdots + C_{s_i})(C_{s_{j-1}+1} + \cdots + C_{s_j})$$
$$= \sum C_{s_{i-1}+j}C_{s_{j-1}+t} = \sum C_{s_{i-1}+\gamma,s_{j-1}+t_1p}C_p$$

i.e.

$$b_{ijp} = \sum_{\gamma,t} C_{s_{i-1}+\gamma,s_{j-1}+t_1p} .$$

In Δ_A put $a_1 = a_2 = \cdots = a_{s_1}$; $a_{s_1+1} = \cdots = a_{s_2-1}$ and so on, i.e. the a's which belong to C's which are in the same class B are put equal to one another. Add

the corresponding rows of Δ'_A then the elements in the row into which the sums are taken are also row elements of Δ_B. The order of $\Delta_B <$ order of Δ'_A. Hence certain columns of Δ_B are repeated in Δ'_A. Subtract the equal columns from one another and we reach a determinant in which Δ_B appears in one corner flanked by a zero matrix. Hence $| \Delta'_A | = | \Delta_B | \varphi$.

3. Factorization of the m-ary p-ic

The characteristic Δ_x is linear in the indeterminates x_i, integral in the c_{ijk} and satisfies the characteristic equation and the reduced characteristic equations which are homogeneous in the x_i's and integral in the c_{ijk}.

Definition 1. A matrix in which the elements are linear in the m-indeterminates x_i and rational or integral in the coefficients of an m-ary p-ic over a field K of characteristic $\neq 2$ will be termed a *linear matrix* over $K(x_1, x_2, \cdots, x_m)$.

We proceed to factorize any m-ary p-ic over K as a product of p mutually commutative linear matrices.

Theorem 10. *An m-ary quadratic can be expressed as the product of commutative linear matrices of order $\leqq 2^{m-1}$.*

Proceed by induction:

$$ax^2 + bxy + cdy^2 = \begin{bmatrix} ax + by, & cy \\ -dy\ , & x \end{bmatrix}\begin{bmatrix} x\ , & -cy \\ dy, & ax + by \end{bmatrix} \tag{7}$$

$$\begin{aligned} ax^2 + by^2 + cz^2 + dyz + ezx + fxy & \\ &= ax^2 + x(fy + ez) + by^2 + dyz + cz^2 \\ &= ax^2 + x(fy + ez) + C \cdot D \\ &= \begin{bmatrix} ax + fy + ez, & c \\ -D\quad , & x \end{bmatrix}\begin{bmatrix} x, -C \\ D, ax + fy + ez \end{bmatrix} \end{aligned} \tag{8}$$

where

$$C = \begin{bmatrix} by + dz, & cz \\ -z\ , & y \end{bmatrix}, \qquad D = \begin{bmatrix} y, & -cz \\ z, & by + dz \end{bmatrix}$$

and where in (8) x, $ax + fy + ez$ are scalar matrices of order 2, i.e.

$$= \begin{bmatrix} ax + fy + ez, & 0\ , & by + dz, & cz \\ 0\ , & ax + fy + ez, & -z\ , & y \\ -y\ , & cz\ , & x\ , & 0 \\ -z\ , & -by - dz\ , & 0\ , & x \end{bmatrix}$$

$$\times \begin{bmatrix} x, & 0\ , & -by - dz\ , & -cz \\ 0, & x\ , & z\ , & -y \\ y, & -cz\ , & ax + fy + cz, & 0 \\ z, & by + dz, & 0\ , & ax + fy + ez \end{bmatrix}.$$

Assume that the $(m-1)$-ary quadratic can be expressed in commutative linear matrix factors of order 2^{m-2}. Then the m-ary quadratic can be written $f = ax^2 + bx + C - D$ where b is linear in $x_2, x_3, \cdots, x_m$ and C and D are commutative linear matrices of order 2^{m-2}. Hence

$$f = \begin{bmatrix} ax+b, & C \\ -D, & x \end{bmatrix}\begin{bmatrix} x, & -C \\ D, & ax+b \end{bmatrix}$$

is factorized in linear commutative matrices of order 2^{m-1}. Hence the theorem may be proved by induction.

In special cases the matrix order may be less than 2^{m-1} and if the restriction as to rationality be abandoned the matrix order may also be reduced further.

Corollary 1. *Similarly the bilinear form* $\sum a_{ij}x_iy_j$ *may be factorized in commutative linear matrices of order* $\leqq 2^m$.

Corollary 2. *A linear form* $\sum a_ix_i$ *may be factorized in commutative linear matrices of order* 2^{2m} *in which the elements are either* $\pm x_i$ *or* $\pm a_i$ *or zero.*

Corollary 3. *The determinant of any matrix factor* $= f^t$, $t = 2^{m-1}$.

Corollary 4. *The factor matrices satisfy the same reduced and the same characteristic equations.*

A *tower* of matrix rings will now be constructed in which the m-ary p-ic may be factorized in p commutative linear matrices.

Let $f = A_1A_2 \cdots A_\gamma + B_1B_2 \cdots B_s$ be scalar in $K(x_1, x_2, \cdots, x_m)$ and let $A_1, A_2, \cdots, A_\gamma$ be mutually commutative linear matrices of the same order μ and $B_1, B_2, \cdots, B_s$ be also mutually commutative linear matrices and of the same order ν. Let $A = \sum \alpha_iA_i$ and $B = \sum \beta_jB_j$, α_i, β_j being indeterminates. In the direct product the matrix rings in which A and B lie let $A \to A'$ where A is repeated ν times in the leading diagonal and let $B \to B'$ in which each element of B is repeated μ times as a scalar matrix. Then $A'B' = B'A' = A \times B$. Hence in the total matrix ring of order $\mu\nu$ every A' is commutative with every B_j' as well as with the remaining A_k''s. Also $(A_iA_j)' = A_i'A_j'$ and $(B_iB_j)' = B_i'B_j'$. Further A_i', B_j' satisfy the same reduced equations as A_i and B_j respectively. Suppose also that $A_1A_2 \cdots A_\gamma$ and $B_1B_2 \cdots B_s$ are each scalars in $K(x_1, x_2, \cdots, x_m)$ then so are $A_1'A_2' \cdots A_\gamma'$ and $B_1'B_2' \cdots B_s'$ in the new matrix ring. Hence $A_1'A_2' \cdots A_\gamma' + B_1'B_2' \cdots B_s'$ is a scalar matrix of order $\mu\nu$ in which f is repeated $\mu\nu$ times in the leading diagonal. In general $A_i'A_j' + B_l'B_m' \neq (A_iA_j + B_lB_m)'$.

We proceed to prove by induction the principal theorem:

Theorem 11. *An m-ary p-ic, f, over a commutative ring K can be expressed as the product of p commutative linear matrices of the same order each linear and integral in the coefficients and, if regular in one of the indeterminates, having $f = 0$ as its reduced equation.*

Assume first that f is regular in x, i.e. $f = x^p + a_1x^{p-1} + a_2x^{p-2} + \cdots + a_p$ where a_i, $i = 1, 2, \cdots, p$ is of degree i and is homogeneous in $x_2, x_3, \cdots, x_p$ the coefficients being in the commutative ring K. For the purpose of factoriza-

tion it is unnecessary to take the coefficient of x^p to be 1 but it simplifies the discussion to assume this. Then

$$f = x[x^{p-1} + a_1x^{p-2} + \cdots + a_{p-1}] + a_p .$$

Assume that it has been proved that both the m-ary p-ic a_p and the m-ary $(p - 1)$-ic [ ] can both be expressed as products of commutative linear matrices viz:

$$f = x(x - x_2)(x - x_3)\cdots(x - x_p) + A_1A_2\cdots A_p . \tag{9}$$

Write these as matrices x_i', A_i' in the direct product M of the matrix rings in which they lie so that, as above, x_i' and A_j' are mutually commutative and for convenience ignore the $'$, then in (9) we may regard the A's and x's as mutually commutative in the matrix ring $K(M)$. Then $x - u$ is a linear matrix factor of f where

$$\begin{bmatrix} x_2, & -A_1, & 0, 0, & \cdots & 0, & 0 \\ 0, & x_3, & -A_2, 0, & \cdots & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & 0, 0, & \cdots & x_p, & -A_{p-1} \\ -A_p, & 0, & 0, 0, & \cdots & 0, & 0 \end{bmatrix}. \tag{10}$$

The remaining factor, found by ordinary division in $K(M)$, is the homogeneous $(p - 1)$-ic in $x, x_2, \cdots, x_m, u$, viz. the $(m + 1)$-ary $(p - 1)$-ic in $K(M, x, x_2, \cdots, x_m, u)$

$$\begin{aligned} &x^{p-1} + (a_1 + u)x^{p-2} + (u^2 + a_1u + a_2)x^{p-3} \\ &\qquad + \cdots + (u^{p-1} + a_1u^{p-2} + \cdots + a_{p-1}). \end{aligned} \tag{11}$$

Hence if we also assume that the $(m + 1)$-ary $(p - 1)$-ic can be factorized as desired then so can the m-ary p-ic. The theorem is true for all the $(m + 1)$-ary quadratic. Hence to complete the induction it is necessary to prove it for the binary p-ic.

The binary p-ic may be factorized in commutative matrices of order $\leqq p!$ of the Galois group of $f = 0$ but such matrices are non-linear in the coefficients, their order not being large enough for our purpose.[3]

The m-ary cubic may be written as $f = x(m\text{-ary quadratic}) + (m - 1)\text{-ary}$ cubic.

Assume that the theorem is true for the $(m - 1)$-ary cubic. Then, since it is true by theorem 10 for the m-ary quadratic, $f = x(x - x_2)(x = x_3) + A_1A_2A_3$. Write this in $K(M)$ as $f' = x(x - x_2')(x - x_3') + A_1'A_2'A_3'$ where $x_2' + x_3'$, $x_2'x_3'$ and $A_1'A_2'A_3'$ are scalars viz. $x_2 + x_3$, x_2x_3, $A_1A_2A_3$ respectively. Then $x - u$ is a linear factor

$$u = \begin{bmatrix} x_2' , & -A_1' , & 0 \\ 0 , & x_3' , & -A_2' \\ -A_3' , & 0 , & 0 \end{bmatrix}$$

[3] A. R. Richardson. Quart. Journal Math., 1934–35, pp. 256–270.

The remaining factor is

$$x^2 + x[u - (x_2' + x_3')] + [u^2 - u(x_2' + x_3') + x_2'x_3']$$

an $(m + 1)$-ary quadratic in $x, x_2, x_3, \cdots, x_m, u$ which can be factorized as desired. The theorem is true for the binary cubic; therefore, by induction, it is true for the m-ary cubic.

The actual factors are:

$$\begin{bmatrix} x - u, & 0 \\ 0 \;, & x - u \end{bmatrix}\begin{bmatrix} x + u - (x_2' + x_3'), & u - x_2' \\ -u + x_3' \;, & x \end{bmatrix}$$

$$\begin{bmatrix} x \;, & -u + x_2' \\ u - x_3', & x + u - (x_2' + x_3') \end{bmatrix}$$

where $x + u - (x_2' + x_3')$ is interpreted as

$$\begin{bmatrix} x - x_3', & -A_1', & 0 \\ 0 \;, & x - x_2', & -A_2' \\ -A_3' \;, & 0 \;, & x - (x_2' + x_3') \end{bmatrix}$$

and $u - x_2'$ as

$$\begin{bmatrix} 0 \;, & -A_1', & 0 \\ 0 \;, & x_3' - x_2', & -A_2' \\ -A_3', & 0 \;, & -x_2' \end{bmatrix}$$

Hence

$$f = \left[\begin{array}{ccc|ccc} x - x_2, & A_1, & 0 & & & \\ 0, & x - x_3, & A_2 & & & \\ A_3, & 0, & x & & & \\ \hline & & & x - x_2, & A_1, & 0 \\ & & & 0, & x - x_3, & A_2 \\ & & & A_3, & 0, & x \end{array}\right]$$

$$\times \left[\begin{array}{ccc|ccc} x - x_3, & -A_1, & 0 & 0, & -A_1, & 0 \\ 0, & x - x_2, & -A_2 & 0, & x_3 - x_2, & -A_2 \\ -A_3, & 0, & x - (x_2 + x_3) & -A_3, & 0, & -x_2 \\ \hline -x_2 + x_3, & A_1, & 0 & x, & 0, & 0 \\ 0, & 0, & A_2 & 0, & x, & 0 \\ A_3, & 0, & x_3 & 0, & 0, & x \end{array}\right]$$

$$\times \left[\begin{array}{ccc|ccc} x, & 0, & 0 & 0, & A_1, & 0 \\ 0, & x, & 0 & 0, & x_2 - x_3, & A_2 \\ 0, & 0, & x & x - x_3, & -A_1, & 0 \\ \hline x_2 - x_3, & -A_1, & 0 & x - x_3, & -A_1, & 0 \\ 0, & 0, & -A_2 & 0, & x - x_2, & -A_2 \\ -A_3, & 0, & -x_3 & -A_3, & 0, & x - (x_2 + x_3) \end{array}\right]$$

where ′ has been dropped. It will be noted that each matrix *zero* is semi-reduced, that the first and the product of the second and third are completely reduced and that the product of the three factors is the scalar matrix having f in each element of the leading diagonal and of order $3!\mu\nu$.

The binary quadratic depends for its factorization on that of a ternary cubic which, as above, can be factorized. Hence the theorem is true for the binary quadratic. *The binary p-ic* depends for its factorization on that of the ternary $(p - 1)$-ic only since the term $a_p y^p$ is already in linear commutative factors. Hence, by induction, the theorem is true for the binary p-ic and, by further induction, for the m-ary p-ic.

COROLLARY 1. *The theorem holds for a multilinear form.*

COROLLARY 2. *Since the matrix factors are linear in the coefficients of the forms, $F = \sum \lambda_i f_i$ will have factors linear in the λ_i's. Therefore any set of m-ary p-ics in the same indeterminates may be factorized simultaneously in the same matrix ring.*

COROLLARY 3. *There are minimum orders to the matrix rings in which a special m-ary p-ic or in which every m-ary p-ic over K can be factorized linearly.*

COROLLARY 4. *The factorization is not unique but since the invariant factors and reduced equations are the same, two factorizations of the same order are similar.*

COROLLARY 5. *f may be regarded as a form in $m + p'$ indeterminates if the p' coefficients are themselves regarded as indeterminate. Hence f may be factorized in matrices in which the elements are sums of the coefficients and indeterminates each multiplied by $+1$, -1 or 0. Such a factorization will be termed complete.*

Factorization of irregular f. The preceding factorization applies to regular f's of which the characteristic and reduced equations of the class-ring are specimens. An irregular f may be factorized thus: arrange the terms of f in any order and consider two of them $\alpha_1\alpha_2 \cdots \alpha_p + \beta_1\beta_2 \cdots \beta_p$ where the α's and β's need not all be different. Then

$$u = \begin{bmatrix} \alpha_1, & \beta_1, & 0, \cdots, & 0 \\ 0, & \alpha_2, & \beta_2, \cdots, & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \beta_p, & 0, & 0, \cdots, & \alpha_p \end{bmatrix}$$

satisfies the equation

$$\prod (\lambda - \alpha_i) + (-1)^p \prod \beta_i = 0. \tag{12}$$

The conjugates of u may be calculated thus: divide by $\lambda - u$ obtaining a $(p - 1)$-ic regular in λ. This may be factorized as in (11). Then u and its conjugates are linear commutative matrices such that

$$u u_2 \cdots u_p = \alpha_1\alpha_2 \cdots \alpha_p + \beta_1\beta_2 \cdots \beta_p .$$

Repeat the process with a further term $\gamma_1\gamma_2 \cdots \gamma_p$ of f. The γ's are commutative with the u's and the above method leads to a set of conjugate commutative linear matrices $\omega\omega_2 \cdots \omega_p = \alpha_1\alpha_2 \cdots \alpha_p + \beta_1\beta_2 \cdots \beta_p + \gamma_1 \cdots \gamma_p$. By successive applications of the process, f itself may be factorized as desired. This is the practical method of procedure.

The matrix order in which the factorization takes place may be reduced if the restriction as to rationality in the coefficients is abandoned. Thus the quaternary cubic may be expressed *irrationally* as

$$f = x_1^3 + x_2^3 + x_3^3 + x_4^3 + x_5^3$$

$$x_2^3 + x_3^3 = \begin{bmatrix} x_2 + x_3, & 0 \\ 0, & x_2 + x_3 \end{bmatrix}\begin{bmatrix} x_2 - x_3, & x_3 \\ -x_3, & x_2 \end{bmatrix}\begin{bmatrix} x_2, & -x_3 \\ x_3, & x_2 - x_3 \end{bmatrix} = ABC$$

$$x_4^3 + x_5^3 = \begin{bmatrix} x_4 + x_5, & 0 \\ 0, & x_4 + x_5 \end{bmatrix}\begin{bmatrix} x_4 - x_5, & x_5 \\ -x_5, & x_4 \end{bmatrix}\begin{bmatrix} x_4, & -x_5 \\ x_5, & x_4 - x_5 \end{bmatrix} = PQR$$

Let

$$u = \begin{bmatrix} A & P & 0 \\ 0 & B & Q \\ R & 0 & C \end{bmatrix}$$

$$S\cdot T = \begin{bmatrix} u - (A + B + C), & 0, & B + C, & B \\ 0, & u - (A + B + C), & -C, & A \\ -A, & B, & u, & 0 \\ -C, & -B - C, & 0, & u \end{bmatrix}$$

$$\times \begin{bmatrix} u, & 0, & -B - C, & -B \\ 0, & u, & C, & -A \\ A, & -B, & u - (A + B + C), & 0 \\ C, & B + C, & 0, & u - (A + B + C) \end{bmatrix}$$

Then:

$$\begin{bmatrix} u, & 0 \\ 0, & u \end{bmatrix}\begin{bmatrix} u - (A + B + C), & S \\ -T, & 0 \end{bmatrix}\begin{bmatrix} 0, & -S \\ T, & u - (A + B + C) \end{bmatrix} = \bar{u}\bar{v}\bar{w}$$

for A, B, C, P, Q, R are mutually commutative.

$$f = x_1^3 + \bar{u}\bar{v}\bar{w} = \begin{bmatrix} x_1 - w, & 0 \\ 0, & x_1 - w \end{bmatrix}\begin{bmatrix} x_1 + w, & -w \\ w, & x_1 \end{bmatrix}\begin{bmatrix} x_1, & w \\ -w, & x_1 + w \end{bmatrix}$$

where

$$w = \begin{bmatrix} 0, & -\bar{u}, & 0 \\ 0, & 0, & -\bar{v} \\ -\bar{w}, & 0, & 0 \end{bmatrix}$$

Hence f may be factorized in matrices of order 144 which is the same as that for the ternary cubic having rational factors. If the ternary cubic is expressed irrationally as $x^3 + y^3 + z^3 + bxyz$ it can be factorized in linear matrices of order 48.

If f is regular in x then the reduced equation for each matrix factor is $f = 0$. Let

$$u = \begin{bmatrix} A_1, & C_1, & 0, & \cdots, & 0, & 0 \\ 0, & A_2, & C_2, & \cdots, & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & 0, & \cdots, & A_{p-1}, & C_{p-1} \\ C_p, & 0, & 0, & \cdots, & 0, & A_p \end{bmatrix} \tag{13}$$

where A_i, C_j; $i, j = 1, 2, \cdots, p$ are mutually commutative matrices each of order μ. Let $T = (t_{i,j})$ where i, j denote row and column respectively and

$$t_{i,j} = (-1)^{p-i} c_1 c_2 \cdots c_{i-1} c_p h_{p-(i+j)}(A_1, A_2, \cdots, A_j, A_p),$$

$$j < p, \qquad t_{i,p} = (-A_p)^{p-_1}$$

where $h_s(\alpha_1, \alpha_2 \cdots \alpha_j)$ is the sum of the homogeneous products of $\alpha_1, \alpha_2, \cdots \alpha_j$ taken s at a time and $h_0 = 1$, $h_s = 0$. Then

$$TuT^{-1} = u' = \begin{bmatrix} a_1, & a_2, & \cdots, & a_{p-1}, & a_p \\ -1, & 0, & \cdots, & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots, & -1, & 0 \end{bmatrix} \tag{14}$$

where $a_p = A_1 A_2 \cdots A_p + (-1)^{p-1} c_1 c_2 \cdots c_p$ and a_i is the elementary symmetric function, $\sum A_1 A_2 \cdots A_i$ of $A_1, A_2, \cdots, A_p$ of degree i. Hence u' satisfies $f = 0$. In general T is non-singuar in $K(M)$ since $|T| = t_{p-1,1}\, t_{p-2,2} \cdots t_{p-i,i} t_{1,p-1} = (-1)^{\nu} c_1^{p-2} c_2^{p-3} \cdots c_{p-2} c_p^{p-1}$ where $\nu = 2$ or 1 according as $p \equiv 0$ or 1; or 2 or 3 (mod 4).

Hence unless all the C's are singular, T is non-singular in $K(M)$ and this suffices for our present purpose, for in the m-ary p-ic $a_p = A_1 A_2 \cdots A_p \neq 0$.

Each element in (14) is a scalar matrix of order μ. Hence u' is similar to u and can be transformed into the matrix in which the canonical elementary matrix in K

$$u'' = \begin{bmatrix} a_1, & a_2, & \cdots & , a_p \\ -1, & 0, & \cdots & , 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots -1, & 0 \end{bmatrix}$$

is repeated μ times in the leading diagonal. Hence u satisfies the same reduced equation in K as u'' viz. $f = 0$.

The $(p - 1)$ conjugates of u also satisfy $f = 0$ as reduced equation in $K(M)$. Thus u_2 satisfies (11) in $K(M, u)$ and since u_2 is equivalent to an elementary canonical matrix in $K(M, u)$, (11) is its reduced equation. Hence, in $K(M)$, $f = 0$ is its reduced equation. A repetition of the argument shows that u_3 is equivalent to an elementary canonical matrix having a reduced equation in $K(M, u, u_2)$ of degree $(p - 1)$ and hence having (11) as reduced equation in $K(M, u)$, i.e. $f = 0$ in $K(M)$. Similarly the other conjugates have $f = 0$ as reduced equation.

These conjugates of u can be transformed into (14). Actual transforming matrices for cubics and quadratics are:

$$T = \begin{bmatrix} x_3^2, & A_1(x_2 + x_3), & 0, & 0, & A_1 x_2, & A_1 A_2 \\ -x_3, & -A_1, & 0, & 0, & -A_1, & 0 \\ 1, & 0, & 0, & 0, & 0, & 0 \\ -2A_3 x_3, & -A_1 A_3, & -x_3(x_2 + x_3), & -A_1 x_3, & -A_1 x_3, & -x_2 x_3 \\ A_3, & 0, & x_3, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0, & 0, & 1 \end{bmatrix}$$

transforms

$$\begin{bmatrix} \bar{u}, & 0 \\ 0, & \bar{u} \end{bmatrix}$$

into the second zero where

$$\bar{u} = \begin{bmatrix} -1, & a_2\,, & a_3 \\ a_1\,, & 0\,, & 0 \\ 0\,, & -1, & 0 \end{bmatrix}$$

In general T is non-singular, e.g. for the binary cubic

$$|\,T\,| = a_3(a_3^2 - a_1a_2).$$

For the quadratic

$$T = \begin{bmatrix} T_{11}\,, & T_{12}\,, & T_{13} \\ T_{21}\,, & T_{22}\,, & T_{23} \\ T_{31}\,, & T_{32}\,, & T_{33} \end{bmatrix}$$

where

$$T_{11} = \begin{bmatrix} -a_1a_2 + a_3\,, & a_2(a_1^2 - a_2), & -a_3(a_1^2 - a_2), & a_4(a_1^2 - a_2) \\ -a_2 \quad, & a_1a_2 \quad, & -a_1a_3 \quad, & a_1a_4 \\ 0 \quad, & a_2 \quad, & -a_3 \quad, & a_4 \\ 1 \quad, & 0 \quad, & 0 \quad, & 0 \end{bmatrix}$$

$$T_{12} = \begin{bmatrix} a_1a_3 - a_4\,, & a_2a_3 - a_1^2a_2\,, & a_4(a_1^2 - a_2), & 0 \\ a_3 \quad, & -a_1a_3 \quad, & a_1a_4 \quad, & 0 \\ 0 \quad, & -a_3 \quad, & a_4 \quad, & 0 \\ 0 \quad, & 0 \quad, & 0 \quad, & 0 \end{bmatrix}$$

$$T_{13} = \begin{bmatrix} -a_1a_4\,, & a_4(a_1^2 - a_2), & 0, & 0 \\ -a_4 \;, & a_1a_4 \quad, & 0, & 0 \\ 0 \;, & a_4 \quad, & 0, & 0 \\ 0 \;, & 0 \quad, & 0, & 0 \end{bmatrix}$$

$$T_{21} = \begin{bmatrix} 0, & -1, & a_1\,, & 0 \\ 0, & 0\,, & 1\,, & 0 \\ 0, & 0\,, & 0\,, & 0 \\ 0, & 0\,, & 0\,, & 0 \end{bmatrix}, \quad T_{22} = \begin{bmatrix} -1, & a_1\,, & -a_2\,, & 0 \\ 0\,, & 0\,, & 0\,, & 0 \\ 0\,, & 0\,, & 1\,, & 0 \\ 0\,, & 0\,, & 0\,, & 0 \end{bmatrix}, \quad T_{23} = \begin{bmatrix} 0, & 0, & 0, & a_4 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0 \end{bmatrix},$$

$$T_{31} = \begin{bmatrix} 0, & 0, & -1, & a_1 \\ 0, & 0, & 0\,, & 1 \\ 0, & 0, & 0\,, & 0 \\ 0, & 0, & 0\,, & 0 \end{bmatrix}, \quad T_{32} = \begin{bmatrix} 0, & -1, & a_1\,, & -a_2 \\ 0, & 0\,, & 0\,, & 0 \\ 0, & 0\,, & 0\,, & 1 \\ 0, & 0\,, & 0\,, & 0 \end{bmatrix}, \quad T_{33} = \begin{bmatrix} -1, & a_1\,, & -a_2\,, & a_3 \\ 0\,, & 0\,, & 0\,, & 0 \\ 0\,, & 0\,, & 0\,, & 0 \\ 0\,, & 0\,, & 0\,, & 1 \end{bmatrix}$$

transforms

$$\begin{bmatrix} u, & 0, & 0 \\ 0, & u, & 0 \\ 0, & 0, & u \end{bmatrix}$$

where

$$u = \begin{bmatrix} -a_1, & a_2, & -a_3, & a_4 \\ -1, & 0, & 0, & 0 \\ 0, & -1, & 0, & 0 \\ 0, & 0, & -1, & 0 \end{bmatrix}$$

into the second zero, viz:

$$\begin{bmatrix} -u - a_1, & u^2 + a_1 u + a_2, & -(u^3 + a_1 u^2 + a_2 u + a_3) \\ -1, & 0, & 0 \\ 0, & -1, & 0 \end{bmatrix}.$$

A result which is useful when calculating the conjugate zeros of u is: if

$$u = \begin{bmatrix} -a_1, & a_2, & -a_3, & \cdots, & (-1)^p a_p \\ -1, & 0, & 0, & \cdots, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots & -1, & 0 \end{bmatrix}$$

then

$$u^s + a_1 u^{s-1} + a_2 u^{s-2} + \cdots + a_s$$

$$= \left[\begin{matrix} 0, & a_{s+1}, & -a_{s+2}, & \cdots & \cdots \\ 0, & 0, & a_{s+1}, & -a_{s+2}, & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & (-1)^{s-2} a_2, & \cdots & \cdots \\ (-1)^s, & (-1)^{s-1} a_1, & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots & \cdots & \cdots \end{matrix} \right.$$

$$\left. \begin{matrix} +(-1)^s a_p, & 0, & 0, & \cdots & 0, & 0 \\ +(-1)^s a_{p-1}, & (-1)^s a_p, & 0, & \cdots & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & a_{s+1}, & -a_{s+2}, & \cdots & (-1)^{s-1} a_{p-1}, & (-1)^s a_p \\ a_s, & 0, & 0, & \cdots & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (-1)^s, & (-1)^{s-1} a_1, & \cdots & \cdots & \cdots & a_s \end{matrix} \right]$$

and in particular

$$u^{p-1} + a_1 u^{p-2} + \cdots + a_{p-1} = (u - x_2)(u - x_3) \cdots (u - x_p)$$

$$= \begin{bmatrix} 0, & a_p, & 0, & \cdots & 0 \\ 0, & 0, & a_p, & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots & \cdots & a_p \\ (-1)^{p-1}, & (-1)^{p-2} a_1, & \cdots & \cdots & a_{p-1} \end{bmatrix}$$

4. Pseudo-representations

The characteristic polynomial of the class-ring may now be factorized in linear commutative matrix factors

$$|\lambda - \Delta_x| = \prod_j [\lambda - \sum x_i X_{i,j}]$$

$X_{i,j}$ being the matrix coefficient of x_i in the j^{th} linear factor. The number of factors is equal to the number of classes and each factor gives rise to a matrix-ring in which the correspondence $C_i \rightarrow X_{i,j}$ can be established.

DEFINITION. *$X_{i,j}$ will be termed a representative of C_i in the j^{th} matrix-ring.*

DEFINITION. *The p matrix-rings will be termed conjugate rings.*

In group theory the correspondence $C_i \rightarrow X_{i,j}/X_{0,j}$ is a representation, but in general this is not so. Even when $C_i \rightarrow X_{i,1}$ gives a representation, the conjugate matrices $X_{i,j}$ may not do so, although the general number in the j^{th} matrix ring is similar to the corresponding general number in the k^{th} ring, i.e. although $T_{j,k}$ exists such that $T_{j,k}^{-1} (\sum x_i X_{i,j})\ T_{j,k} = \sum x_i X_{i,k}$, $T_{j,k}$ depends on x_i's and does not transform $X_{i,j}$ into $X_{i,k}$. There may however exist a *pseudo-representation* in which if $C_i C_j = \sum C_{ijp} C_p$ then

$$\prod_k [X_{i,k}\, X_{j,k} - \sum c_{ijp}\, X_{p,k}] = 0, \tag{15}$$

although the factors in this product are not necessarily conjugates.

Pseudo-representation of the total matrix algebra of two rowed matrices. Let $e_{11}, e_{12}, e_{21}, e_{22}$ denote $\begin{pmatrix}1, & 0\\0, & 0\end{pmatrix}, \begin{pmatrix}0, & 1\\0, & 0\end{pmatrix}, \begin{pmatrix}0, & 0\\1, & 0\end{pmatrix}, \begin{pmatrix}0, & 0\\0, & 1\end{pmatrix}$ respectively. The characteristic polynomial of the general number $x = \alpha e_{11} + \beta e_{12} + \gamma e_{21} + \delta e_{22}$ is

$$x^2 - x(\alpha + \delta) + \alpha\delta - \beta\gamma.$$

One factorization of this is:

$$\begin{bmatrix} x - (\alpha+\delta), & 0, & \alpha, & \beta \\ 0, & x - (\alpha+\delta), & \gamma, & \delta \\ -\delta, & \beta, & x, & 0 \\ \gamma, & -\alpha, & 0, & x \end{bmatrix} \begin{bmatrix} x, & 0, & -\alpha, & -\beta \\ 0, & x, & -\gamma, & -\delta \\ \delta, & -\beta, & x - (\alpha+\beta), & 0 \\ -\gamma, & \alpha, & 0, & x - (\alpha+\beta) \end{bmatrix}$$

which gives rise to two sets of matrices $X_{i,j}$ corresponding respectively to e_{11}, e_{12}, e_{21}, e_{22} viz.:

$$\begin{bmatrix} 1, & 0, & -1, & 0 \\ 0, & 1, & 0, & 0 \\ 0, & 0, & 0, & 0 \\ 0, & 1, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 0, & -1 \\ 0, & 0, & 0, & 0 \\ 0, & -1, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 0, & 0 \\ 0, & 0, & -1, & 0 \\ 0, & 0, & 0, & 0 \\ -1, & 0, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 1, & 0, & 0, & 0 \\ 0, & 1, & 0, & -1 \\ 1, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix} \tag{16}$$

$$(17)\quad \begin{bmatrix} 0, & 0\ , & 1, & 0 \\ 0, & 0\ , & 0, & 0 \\ 0, & 0\ , & 1, & 0 \\ 0, & -1, & 0, & 1 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 0, & 1 \\ 0, & 0, & 0, & 0 \\ 0, & 1, & 0, & 0 \\ 0, & 0, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 0, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 0 \\ 1, & 0, & 0, & 0 \end{bmatrix}, \begin{bmatrix} 0\ , & 0, & 0, & 0 \\ 0\ , & 0, & 0, & 1 \\ -1, & 0, & 1, & 0 \\ 0\ , & 0, & 0, & 1 \end{bmatrix}$$

These give a pseudo-representation for, using (16)

$$Z_1 = [xx' - \{(\alpha\alpha' + \beta\gamma')e_{11} + (\alpha\beta' + \beta\delta')e_{12} + \gamma\alpha' + \delta\gamma')e_{21} + (\gamma\beta' + \delta\delta')e_{22}\}]$$

$$= \begin{bmatrix} \delta\alpha' - \gamma\beta', & \alpha\beta' - \beta\alpha', & \beta\gamma' - \delta\alpha', & \beta\delta' - \delta\beta' \\ \delta\gamma' - \gamma\delta', & \alpha\delta' - \beta\gamma', & \gamma\alpha' - \alpha\gamma', & \gamma\beta' - \alpha\delta' \\ \delta\alpha' - \gamma\beta', & \alpha\beta' - \beta\alpha', & \beta\gamma' - \delta\alpha', & \beta\delta' - \delta\beta' \\ \delta\gamma' - \gamma\delta', & \alpha\delta' - \beta\gamma', & \gamma\alpha' - \alpha\gamma', & \gamma\beta' - \alpha\delta' \end{bmatrix}$$

and, using (17)

$$Z_2 = \begin{bmatrix} \beta\gamma' - \alpha\delta', & \alpha\beta' - \alpha'\beta, & \alpha\delta' - \beta\gamma', & \beta\alpha' - \alpha\beta' \\ \delta\gamma' - \gamma\delta', & \gamma\beta' - \delta\alpha', & \gamma\delta' - \delta\gamma', & \delta\alpha' - \gamma\beta' \\ \gamma\beta' - \alpha\delta', & \delta\beta' - \beta\delta', & \alpha\delta' - \gamma\beta', & \beta\delta' - \delta\beta' \\ \alpha\gamma' - \gamma\alpha', & \beta\gamma' - \delta\alpha', & \gamma\alpha' - \alpha\gamma', & \delta\alpha' - \beta\gamma' \end{bmatrix}$$

and $Z_1Z_2 = 0$, i.e. the matrices (16) and (17) give a pseudo-representation of the total matrix algebra. Nevertheless Z_2 is not a conjugate of Z_1.

There are two other matters which may be mentioned. In the group ring the characters are orthogonal. In the general case this is not so and a biorthogonal relation takes its place. Suppose that the matrix equations

$$\sum_i \alpha_{i,j} X_{i,j} = 1, \qquad \sum_j \alpha_{i,j} X_{i,s} = 0, \qquad s \neq j$$

can be solved for the $\alpha_{i,j}$. This will be so if the determinant $|X_{i,j}|$ regarded as a determinant in K is non-zero. Then the α's and X's are biorthogonal, viz:

$$\sum_j X_{i,j}\alpha_{i,j} = n; \qquad \sum_j X_{i,j}\alpha_{k,j} = 0;$$

$$\sum_i \alpha_{i,j} X_{i,j} = n; \qquad \sum_i \alpha_{i,j} X_{i,k} = 0,$$

where n is an element of K invariant in the class-ring; in the group ring it is the group order.

The matrix ring generated by the α's is not in general the same as that generated by the $X_{i,j}$ but the two rings and the corresponding polynomials are related in a way which has some geometrical significance.

The second matter of interest is that $f = 0$ is the reduced characteristic equation of each of the matrix factors. If $f = 0$ is reducible in K, e.g. $f = f_1 f_2 \cdots f_s$

then the matrices may be constructed, as above, having each of these factors or any combination of them as reduced equations. These remarks will be illustrated by examples.

EXAMPLE 1.

$$\begin{array}{c|cc} & C_1 & C_2 \\ \hline C_1 & C_1 & C_2 \\ C_2 & C_2 & C_1 + C_2 \end{array}$$

The characteristic equation is

$$x^2 - x(2x_1 + x_2) + (x_1^2 + x_1x_2 - x_2^2) = 0 \tag{18}$$

This may be factorized as

$$\begin{bmatrix} x - (x_1 + x_2), & x_2 \\ x_2 \quad , & x - x_1 \end{bmatrix} \begin{bmatrix} x - x_1, & -x_2 \\ -x_2 \quad , & x - (x_1 + x_2) \end{bmatrix}$$

giving the true representations

$$\begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 0, & 1 \\ 1, & 1 \end{pmatrix}; \begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 1\ , & -1 \\ -1, & 0 \end{pmatrix}; = x_{11}, x_{12}; x_{21}, x_{22}$$

respectively. The corresponding $\alpha_{i,\ j}$ are

$$\begin{pmatrix} 3\ , & -1 \\ -1, & 2 \end{pmatrix}, \begin{pmatrix} -1, & 2 \\ 2\ , & 1 \end{pmatrix}; \begin{pmatrix} 2, & 1 \\ 1, & 3 \end{pmatrix}, \begin{pmatrix} 1\ , & -2 \\ -2, & -1 \end{pmatrix}; = \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}$$

respectively. The biorthogonal relations are:

$$\begin{aligned} x_{11}\alpha_{11} + x_{12}\alpha_{12} &= 5; & x_{21}\alpha_{11} + x_{22}\alpha_{12} &= 0; \\ x_{11}\alpha_{21} + x_{12}\alpha_{22} &= 0; & x_{21}\alpha_{21} + x_{22}\alpha_{22} &= 5; \\ \alpha_{11}x_{11} + \alpha_{21}x_{21} &= 5; & \alpha_{11}x_{12} + \alpha_{21}x_{22} &= 0; \\ \alpha_{12}x_{11} + \alpha_{22}x_{21} &= 0; & \alpha_{12}x_{12} + \alpha_{22}x_{22} &= 5. \end{aligned}$$

A different factorization of (18)

$$\begin{bmatrix} x - (2x_1 + x_2), & 0 \quad , & x_1 + x_2, & x_2 \\ 0 \quad , & x - (2x_1 + x_2), & x_2 \quad , & x_1 \\ -x_1 \quad , & x_2 \quad , & x \quad , & 0 \\ x_2 \quad , & -x_1 - x_2 \quad , & 0 \quad , & x \end{bmatrix}$$

$$\begin{bmatrix} x\ , & 0\ , & -x_1 - x_2\ , & -x_2 \\ 0\ , & x\ , & -x_2\ , & -x_1 \\ x\ , & -x_2\ , & x - (2x_1 + x_2), & 0 \\ -x_2, & x_1 + x_2, & 0\ , & x - (2x_1 + x_2) \end{bmatrix}$$

gives the pseudo-representations

$$\begin{bmatrix} 2, & 0, & -1, & 0 \\ 0, & 2, & 0\,, & -1 \\ 1, & 0, & 0\,, & 0 \\ 0, & 1, & 0\,, & 0 \end{bmatrix}, \quad \begin{bmatrix} 1\,, & 0\,, & -1, & -1 \\ 0\,, & 1\,, & -1, & 0 \\ 0\,, & -1, & 0\,, & 0 \\ -1, & 1\,, & 0\,, & 0 \end{bmatrix};$$

$$\begin{bmatrix} 0\,, & 0\,, & 1, & 0 \\ 0\,, & 0\,, & 0, & 1 \\ -1, & 0\,, & 2, & 0 \\ 0\,, & -1, & 0, & 2 \end{bmatrix}, \quad \begin{bmatrix} 0, & 0\,, & 1, & 1 \\ 0, & 0\,, & 1, & 0 \\ 0, & 1\,, & 1, & 0 \\ 1, & -1, & 0, & 1 \end{bmatrix} \equiv x_{11}\,,\ x_{12}\,;\ x_{21}\,,\ x_{22}$$

for which it is readily verified that

$$(x_{12}^2 - x_{11} - x_{12})(x_{22}^2 - x_{21} - x_{22}) = 0; \qquad (x_{11}^2 - x_{11})(x_{21}^2 - x_{21}) = 0$$

$$(x_{11}x_{12} - x_{12})(x_{21}x_{22} - x_{22}) = 0.$$

EXAMPLE 2. *The symmetric group of order* 3.

The reduced equation of the group-ring is

$$[x - (x_1 + 3x_2 + 2x_3)][x - (x_1 - x_3)][x - (x_1 - 3x_2 + 2x_3] = 0$$

Taking the first two factors together and factorizing in matrices of order 2 we get as pseudo-representations

$$\begin{pmatrix} 2, & -1 \\ 1, & 0 \end{pmatrix}, \begin{pmatrix} -3, & 0 \\ -3, & 0 \end{pmatrix}, \begin{pmatrix} 1, & 1 \\ 2, & 0 \end{pmatrix}; \begin{pmatrix} 0\,, & 1 \\ -1, & 2 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 3, & -3 \end{pmatrix}, \begin{pmatrix} 0\,, & -1 \\ -2, & 1 \end{pmatrix}$$

$$= x_{21}\,,\ x_{22}\,,\ x_{23}\,;\ x_{31}\,,\ x_{32}\,,\ x_{33}$$

representatives of the classes (1), (12), (123). Writing the remaining factors as

$$\begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 3, & 0 \\ 0, & 3 \end{pmatrix}, \begin{pmatrix} 2, & 0 \\ 0, & 2 \end{pmatrix} = x_{11}\,,\ x_{12}\,,\ x_{13}$$

the biorthogonal set is:

$$\begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}; \begin{pmatrix} 1\,, & 0 \\ -3, & 4 \end{pmatrix}, \begin{pmatrix} -1, & 0 \\ -1, & 0 \end{pmatrix}, \begin{pmatrix} 1, & 0 \\ 3, & -2 \end{pmatrix};$$

$$\begin{pmatrix} 1, & 3 \\ 0, & 4 \end{pmatrix}, \begin{pmatrix} -1, & 1 \\ 0\,, & 0 \end{pmatrix}, \begin{pmatrix} 1, & -3 \\ 0, & -2 \end{pmatrix}$$

with

$$\sum_j x_{i,j}\alpha_{ij} = 3!, \qquad \sum_j x_{ij}\alpha_{k,j} = 0; \qquad \sum_i \alpha_{ij}x_{ij} = 3!, \qquad \sum_i \alpha_{i,j}x_{i,k} = 0.$$

Similarly any other grouping of the factors of the characteristic equation leads to a pseudo-representation.

The above pseudo-representations generate total matrix rings.

We may also repeat one of the factors, e.g. corresponding to $[x - (x_1 - x_3)]^2$ we have the pseudo-representations

$$\begin{pmatrix} 2, & -1 \\ 1, & 0 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} -2, & 1 \\ -1, & 0 \end{pmatrix}; \begin{pmatrix} 0\,, & 1 \\ -1, & 2 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} 0, & -1 \\ 1, & -2 \end{pmatrix}$$

These do not generate total matrix rings.

EXAMPLE 3. *The symmetric group of order* 4.

Denote the classes (1), (12), (123), (12) (34), (1234) by C_α, C_β, C_γ, C_δ, C_ϵ respectively. Then the characteristic equation has as zeros

$$\alpha + 6\beta + 8\gamma + 3\delta + 6\epsilon; \qquad \alpha - 6\beta + 8\gamma + 3\delta - 6\epsilon; \qquad \alpha - 4\gamma + 3\delta;$$
$$\alpha + 2\beta - \delta - 2\epsilon; \qquad \alpha - 2\beta - \delta + 2\epsilon.$$

Any combination of these gives a pseudo-representation, e.g.
$(\alpha - \delta)^2 - 4(\beta - \epsilon)^2$ gives

$$\begin{pmatrix} 2, & -1 \\ 1, & 0 \end{pmatrix}, \begin{pmatrix} 0\ , & -2 \\ -2, & 0 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} -2, & 1 \\ -1, & 0 \end{pmatrix}, \begin{pmatrix} 0, & 2 \\ 2, & 0 \end{pmatrix};$$
$$\begin{pmatrix} 0\ , & 1 \\ -1, & 2 \end{pmatrix}, \begin{pmatrix} 0, & 2 \\ 2, & 0 \end{pmatrix}, \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix}, \begin{pmatrix} 0, & -1 \\ 1, & -2 \end{pmatrix}, \begin{pmatrix} 0\ , & -2 \\ -2, & 0 \end{pmatrix}.$$

These generate simply isomorphic matrix-rings.

EXAMPLE 4. *The following set has the same representation as the symmetric group of order* 2.

$$\sum : \begin{array}{c|ccc} & a & b & c \\ \hline a & a & c & b \\ b & c & a & a \\ c & b & a & a \end{array}; \qquad A = a; \qquad C_1 = a, \qquad C_2 = b,c \qquad \bmod R_A.$$

Then

$$\begin{array}{c|cc} & C_1 & C_2 \\ \hline C_1 & C_1 & C_2 \\ C_2 & C_2 & 4C_1 \end{array}$$

and the characteristic equation is

$$(x - x_1 + 2x_2)(x - x_1 - 2x_2) = 0.$$

Hence dividing the representatives by the class orders 1, 2, respectively we get 1, 1, 1; 1, -1, -1, as representatives of a, b, c respectively. The class-ring is simply isomorphic with the group-ring generated from the symmetric group of order 2! in which $C_1 \to 1$, $C_2 \to 2(1, 2)$. Therefore the existence of a set of numbers having all the properties of group-characters cannot be taken as evidence of the existence of a group having these as characters.

Conclusion

In addition to the analogy stressed above between the group characteristics and the linear matrix representatives of classes found by factorization of the characteristic polynomial of the class-ring there are many other fields in which the factorization of the m-ary p-ic may be applied.

For example the theory of quadratic forms is known to depend on that of generalized quaternion algebras. This becomes apparent as soon as the form is factorized. There are also applications to the theories of algebraic functions, invariants and arithmetic.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

ON HOMOTOPY TYPE AND DEFORMATION RETRACTS

By R. H. Fox

(Received October 1, 1942)

It has recently been shown by J. H. C. Whitehead[1] that two complexes X and Y belong to the same homotopy type[2] if and only if there is a third complex W of which both X and Y are deformation retracts.[3] I shall show that this theorem holds not merely for complexes but for the most general spaces for which continuity has a meaning. The proof which I give is direct and constructive and avoids the extraneous notions of relative homology and relative homotopy groups which complicate Whitehead's proof.

The concept of homotopy type splits naturally into two concepts which I shall call right- and left-homotopy inversion. In theorems 3.3 and 3.4 I show that right-and-left inversion correspond respectively to deformation and retraction, thus replacing Whitehead's theorem by two "component" theorems. The necessary preliminary study of deformation, retraction, and inversion is carried out in §§1 and 2, and the mapping cylinder, the fundamental tool of our theory, is defined in §3. It should be noted that Whitehead's definition[4] of mapping cylinder is not really satisfactory for the general spaces considered here.

The fundamental theorems of this paper are theorems 3.1 and 3.2. They are generalizations of the theorems (3.3 and 3.4) discussed above. In §4 these fundamental theorems are applied (in another direction) to the Hopf-Pannwitz deformation and also to yield a new characterization of the closure of a homogenous n-dimensional polyhedron.

These theorems (3.1 and 3.2) are of considerable interest in themselves. They exhibit a duality which is quite striking and seem to indicate a relatively unexplored region which I might designate as "algebra of mapping classes". In this connection they should be compared with the fundamental theorem of fibre spaces[5] to which they bear an evident analogy.

In §§5 and 6 certain specializations are considered. They are to be regarded as trends in the following two directions (a) bridging the gap between homotopy type and nucleus[6] (b) bridging the gap between homotopy type and topological type. In §7 I develop an n-dimensional analogue of §3. This is in line with

[1] J. H. C. Whitehead, *Simplicial spaces, nuclei, and m-groups*. Proc. London Math. Soc. 45 (1939), 243–327. The proof referred to is on p. 278.

[2] W. Hurewicz, *Topologie der Deformationen* III. Proc. Akad. Amsterdam 39 (1936), p. 124.

[3] K. Borsuk, *Zur Kombinatorischen Eigenschaften der Retracte*. Fund. Math. 21 (1933), p. 91.

[4] J. H. C. Whitehead, loc. cit., p. 259.

[5] W. Hurewicz and N. E. Steenrod, *Homotopy Relations in Fibre Spaces*, Proc. Nat. Acad. 27 (1941), p. 62, theorem 1.

[6] J. H. C. Whitehead, loc. cit., p. 247.

the viewpoint of my Thesis,[7] especially §§2 and 13; analogous definitions and theorems using homology and n-dimensional homology are quite obvious and are omitted. The n-dimensional homotopy was selected because the n-dimensional homotopy type seems to be related to the so-called $(n + 1)$-group[6] of Whitehead in much the same way that homotopy type is related to nucleus.

1. Deformation and retraction

Mappings[8] f and g of a space A into a space D are said to be *homotopic* (notation: $f \simeq g$) if there is a mapping ξ (called a *homotopy* between f and g) of the product $A \times [0, 1]$ of A with the closed interval $0 \leqq t \leqq 1$ into D such that $\xi_0(a) = f(a)$ and $\xi_1(a) = g(a)$ for every $a \epsilon A$. If A is a subset of D and f is the identity, so that $f(a) = a$, the homotopy ξ is called a *deformation* and the set A is said to be *deformable* in D into $g(A)$. If $D = A$ we say merely that A can be deformed into $g(A)$.

(1.1)[9] *If A can be deformed into B then a mapping f of A into itself is homotopic to the identity if (and only if) $f \mid B$*[10] *is homotopic to the identity.*

If ξ is a deformation of A into B and η is a homotopy between $f \mid B$ and the identity then a homotopy ζ between f and the identity is defined by

$$\begin{aligned} \zeta_t(a) &= \xi_{3t}(a), & 0 \leqq t \leqq \tfrac{1}{3}, \\ &= \eta_{2-3t}(\xi_1(a)), & \tfrac{1}{3} \leqq t \leqq \tfrac{2}{3}, \\ &= f(\xi_{3-3t}(a)), & \tfrac{2}{3} \leqq t \leqq 1. \end{aligned}$$

A subset B of a space A is said to be a *retract* of A if there is a mapping r (called a *retraction*) of A into B such that $r \mid B$ is the identity mapping of B. If r is homotopic to the identity mapping of A then B is called a *deformation retract*[3] of A, r is called a *deformation retraction* and the homotopy is called a *retracting deformation.*

THEOREM 1.2. *In order that B be a deformation retract of A it is necessary and sufficient that B be a retract of A and A be deformable into B.*

This follows from (1.1) by specializing f to be a retraction of A into B.

The example of a point B contained in an n-sphere A ($n \geqq 0$) shows that B may be a retract of A without A being deformable into B. The example of an n-sphere B contained in an $(n + 1)$-cell A ($n \geqq 0$) shows that A may be deformable into B without B being a retract of A. These two statements are equivalent to each other and to the Brouwer fixed point theorem.[11]

---

[7] R. H. Fox, *On the Lusternik-Schnirelmann Category*, Annals of Math. 42 (1941), 333–370.

[8] A mapping of a space A into a space D means a continuous function defined on A with values in D.

[9] This lemma is a trivial generalization of Satz IVa Hilfsatz, Alexandroff and Hopf, *Topologie*, p. 251.

[10] The partial mapping $f \mid B$ is the mapping of B defined by the rule $\{f \mid B\}(b) = f(b)$ for every $b \epsilon B$.

[11] See W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton Mathematical Series 4 (1941) Chapter 5, §1.

(1.3) *If ξ is a deformation of an ANR-set*[12] *A and B is the set of fixed points of ξ_1 then there is a homotopy ζ between $\xi_0 = 1$ and ξ_1 such that the points of B are fixed under each of the mappings ζ_t, $0 \leqq t \leqq 1$.*

Let η be the map of $(A \times [0] + B \times [0, 1] + A \times [1]) \times [0, 1]$ into A defined by

$$\eta_u(a, 0) = a \quad , \quad 0 \leqq u \leqq 1, \qquad a \in A,$$

$$\eta_u(a, t) = \begin{cases} \xi_{t/u}(a), & 0 \leqq t \leqq u \leqq 1, \qquad a \in B, \\ a \quad , & 0 \leqq u \leqq t \leqq 1, \qquad a \in B, \end{cases}$$

$$\eta_u(a, 1) = \xi_1(a) \quad , \quad 0 \leqq u \leqq 1, \qquad a \in A.$$

Since ξ is an extension[13] to $A \times [0, 1]$ of the map η_1, and since A is an ANR-set, it follows that there is an extension[14] ζ to $A \times [0, 1]$ of η_0. This deformation ζ has the required properties.

THEOREM 1.4. *If A and a subset B are ANR-sets then either of the conditions* (i), (ii) *is necessary and sufficient for B to be a deformation retract of A.*

(i) *There is a retracting deformation ξ of A onto B such that the points of B are fixed under each of the mappings ξ_t, $0 \leqq t \leqq 1$.*

(ii) *There is a deformation ξ of A into B such that $\xi_t(B) \subset B$ for every $0 \leqq t \leqq 1$.*

The sufficiency of (i) and the implication (i) $\rightarrow$ (ii) are trivial. That condition (i) is necessary follows from (1.3) by imposing the condition $\xi_1(A) = B$ so that ξ becomes a retracting deformation. If (ii) is assumed then $\xi_1 \mid B$ and $\xi_0 \mid B$ are homotopic in B. Since B is a compact ANR-set and since ξ_1 is an extension of $\xi_1 \mid B$ to A (with values in B) it follows[14] that there is an extension r of $\xi_0 \mid B$ to A. This extension r is clearly a retraction of A onto B. By (1.2) it follows that B is a deformation retract of A.

Appreciation of theorem 1.4 is facilitated by consideration of several examples. The first example is due to Hopf and Pannwitz.[15] B is the pseudomanifold obtained by pinching a meridian of a torus to a point; A is obtained from B by spanning an equator with a 2-cell. A can be deformed into B but condition (ii) is not satisfied (hence B is not a deformation retract and condition (i) is not satisfied either). In the second example A is a bounded portion of the Cartesian plane and B is the set

$$\{0 \leqq x \leqq 1; y = 0\} + \textstyle\sum_{n=1}^{\infty} \{x = 1/n; 0 \leqq y \leqq 1\} + \{x = 0; 0 \leqq y \leqq 1\},$$

hence not an ANR. Condition (ii) is satisfied but B is not a retract of A, hence not a deformation retract of A. In the third example A is the set B of the previous example and B is the point $(0, 1)$. B is a deformation retract of A but

[12] R. H. Fox, *A Characterization of Absolute Neighborhood Retracts*, Bull. Am. Math. Soc. 48 (1942), 271–275.

[13] If f is a mapping of X into Y an extension f^* of f to a space $X^* \supset X$ is a mapping of X^* into Y such that $f^* \mid X = f$.

[14] The Borsuk-Kuratowski theorem: Fox, ibid, p. 273, and Dowker's proof of Borsuk's theorem: Hurewicz and Wallman, ibid, p. 86.

[15] Alexandroff and Hopf, loc. cit. p. 287, fig. 23.

condition (ii) is not satisfied.[16] In the fourth example A is as in the last example and B is the line segment $\{x = 0; 0 \leqq y \leqq 1\}$. Here B is a deformation retract of A. Condition (ii) is satisfied but not[16] condition (i).

2. Homotopy type

Two spaces A and B are said to belong to the same *homotopy type*[2] if there are mappings f of A into B and g of B into A such that the maps gf of A into itself and fg of B into itself are each homotopic to the identity (in A and B respectively). Belonging to the same homotopy type is an equivalence relation.[2]

If mappings f of A into B and g of B into A are such that $gf \simeq 1$ then g will be called a *left homotopy inverse* of f or, briefly, a *left inverse* of f and f will be called a *right inverse* of g. *A two-sided inverse* of f is a mapping which is both a right and left inverse of f. Thus A and B belong to the same homotopy type if and only if there is a mapping of A into B which has a two-sided inverse.

THEOREM 2.[17] *If f has both right and left inverses then it has a 2-sided inverse.*

Let g' and g'' be left and right inverses respectively and let $g = g'fg''$. Then $gf = g'fg''f \simeq g'1f = g'f \simeq 1$ and $fg = fg'fg'' \simeq f1g'' = fg'' \simeq 1$.

It may happen, even for compact ANR-sets A and B, that no map of A into B has a 2-sided inverse although maps of A into B can be found with either right or left inverses. Let O denote the 1-sphere and 8 the figure-eight graph. Let $A = 8 \times 8 \times 8 \times \cdots$ and let $B = O \times 8 \times 8 \times \cdots = O \times A$, so that A and B are compact ANR-sets.[18] The map f' of A into B defined by $f'(t_1, t_2, \cdots) = (p, t_1, t_2, \cdots)$, $p \in O$, has a left inverse g' defined by $g'(t_1, t_2, \cdots) = (t_2, t_3, \cdots)$. The map f'' of A into B defined by $f''(t_1, t_2, t_3, \cdots) = (\alpha(t_1), t_2, t_3, \cdots)$, where α maps 8 into O by folding the top down over the bottom, has a right inverse g'' defined by $g''(t_1, t_2, t_3, \cdots) = (\beta(t_1), t_2, t_3, \cdots)$ where β maps O homeomorphically onto the bottom half of 8. The fundamental group $\pi_1(A)$ of A is the infinite direct product $F_2 \times F_2 \times F_2 \times \cdots$ and the fundamental group $\pi_1(B)$ of B is the direct product $F_1 \times F_2 \times F_2 \times \cdots = F_1 \times \pi_1(A)$, where F_i denotes the free group on i generators ($i = 1, 2$). Since $\pi_1(B)$ has an element $(a, 1, 1, \cdots)$ which commutes with every element of $\pi_1(B)$, and $\pi_1(A)$ has no such element, $\pi_1(A)$ and $\pi_1(B)$ are not isomorphic. Hence[19] no map of A into B can have a 2-sided inverse.

3. Mapping cylinder

For any mapping α of a space[20] M into a space N let $N + C_\alpha$ denote the space obtained from the space N and the cylinder $M \times [0, 1]$ by identifying the point

[16] R. H. Fox, *On the Lusternik-Schnirelmann Category*, Annals of Math. 42 (1941), p. 362.

[17] This proof of the theorem was shown to me by M. M. Day who has proved a theorem on partially ordered sets by exactly the same method.

[18] N. Aronszahn and K. Borsuk, *Sur la somme et le produit combinatoire des rêtractes absolus*. Fund. Math. 18 (1932), Theorem 6, p. 197.

[19] W. Hurewicz, *Topologie der Deformationen* III, Proc. Akad. Amsterdam 39 (1936), p. 125.

[20] Unless otherwise specified spaces considered at the same time are mutually separated.

$n \varepsilon N$ and the closed set $(\alpha^{-1}(n), 1) \varepsilon M \times [1]$. Precisely, $N + C_\alpha$ is the hyperspace[21] of the decomposition of $N + M \times [0, 1]$ into the points (m, t), $0 \leqq t < 1$, of $M \times [0, 1)$ and the (closed) sets $n + (\alpha^{-1}(n), 1)$ of $N + M \times [1]$. Denoting the identification mapping by i, it can easily be proved that $i \mid M$ and $i \mid N$ are homeomorphisms and that $i \mid (f(M) + M \times [0, 1])$ is the identification mapping of the induced decomposition of $f(M) + M \times [0, 1]$. (Note that $i \mid M \times [0, 1]$ is not necessarily the identification mapping of the induced decomposition of $M \times [0, 1]$). Accordingly denote by C_α the hyperspace of the induced decomposition of $f(M) + M \times [0, 1]$ and consider C_α, M and N as subsets of $N + C_\alpha$ so that $i \mid (M + N)$ is the identity mapping. (This justifies the notation $N + C_\alpha$). I shall call $N + C_\alpha$ the mapping cylinder[4] of α; the symbol $\langle m, t\rangle$ will denote the point $i(m, t)$ of C_α, so that $m = \langle m, 0\rangle$ and $\alpha(m) = \langle m, 1\rangle$. If M and N are compact metric then so also is $N + C_\alpha$.[22] The dimension of $N + C_\alpha$ is max $\{\dim M, 1 + \dim N\}$.[23]

THEOREM 3.1. *Let X, Y and Z be topological spaces and let θ be a mapping of X into Z. If f is a mapping of X into Y then there is a mapping g of Y into Z satisfying $gf \simeq \theta$ if and only if θ can be extended*[13] *to $Y + C_f$.*

If θ^* is an extension of θ to $Y + C_f$ let $g = \theta^* \mid Y$. The homotopy ξ defined by

$$\xi_t(x) = \theta^*(\langle x, t\rangle), \qquad x \epsilon X, \qquad 0 \leqq t \leqq 1,$$

is a homotopy between $\xi_0 = \theta^* \mid X = \theta$ and $\xi_1 = \theta^* f = gf$.

Suppose, conversely, that a map g satisfying $gf \simeq \theta$ has been given and that ξ is a homotopy between θ and gf. Then the map θ^* of $Y + C_f$ into Z defined by

$$\theta^*(\langle x, t\rangle) = \xi_t(x), \qquad \langle x, t\rangle \epsilon C_f,$$
$$\theta^*(y) = g(y), \qquad y \epsilon Y,$$

is an extension of θ to $Y + C_f$. $\theta^* \mid C_f$ is continuous because if V is an open set of Z then $f^{-1}(V)$ is an open set of $X \times [0, 1]$ which is the union of sets of the decomposition $\{i^{-1}(\langle x, t\rangle)\}$.

THEOREM 3.2. *Let X, Y, Z, and θ be as in theorem* 3.1. *If g is a mapping of Y into Z then there is a mapping f of X into Y satisfying $gf \simeq \theta$ if and only if θ is homotopic in $Z + C_g$ to a map of X into Y.*

Let ω denote the homotopy in $Z + C_g$ between the identity mapping of Y and the mapping g; explicitly

$$\omega_t(y) = \langle y, t\rangle, \qquad y \epsilon Y, 0 \leqq t \leqq 1.$$

If $\theta \simeq \theta'$ in $Z + C_g$ where $\theta'(X) \subset Y$ then $\theta \simeq \theta' = \omega_0\theta' \simeq \omega_1\theta' = g\theta'$. Thus a mapping f satisfying $gf \simeq \theta$ is the mapping $f = \theta'$.

If, conversely, there is an f such that $gf \simeq \theta$ then $\theta \simeq gf = \omega_1 f \simeq \omega_0 f = f$; thus f is a map of X into Y which is homotopic in $Z + C_g$ to the given map θ.

[21] Alexandroff and Hopf, loc. cit. p. 63.
[22] Ibid., p. 96–99.
[23] Kuratowski, *Topologie* I, p. 127.

On choosing $Z = X$ and $\theta = 1$ in theorem 3.1 we have

THEOREM 3.3. *X is a retract of $Y + C_f$ if and only if f has a left inverse.*

On choosing $Z = X$ and $\theta = 1$ in theorem 3.2 and observing that $Z + C_g$ can be deformed into Z by the homotopy

$$\xi_t(\langle y, s\rangle) = \langle y, s + t(1 - s)\rangle, \qquad \langle x, s\rangle \in C_g,$$

$$\xi_t(z) = z, \qquad z \in Z,$$

we have

THEOREM 3.4.[24] *$Z + C_g$ can be deformed into Y if and only if g has a right inverse.*

Comparison of theorems 3.3 and 3.4 shows a curious "duality" between deformation and retraction.

(3.5) *There is a map f of X into Y such that X is a retract of $Y + C_f$ if and only if there is a map g of Y into X such that $X + C_g$ can be deformed into Y.*

This is a corollary of the more general "duality" implied by 3.1 and 3.2.

(3.6) *There is a map f of X into Y such that θ can be extended to $Y + C_f$ if and only if there is a map g of Y into Z such that θ is homotopic in $Z + C_g$ to a mapping of X into Y.*

From theorems 1.2, 2, 3.3, and 3.4 follows

THEOREM 3.7. *X is a deformation retract of $Y + C_f$ if and only if f has a 2-sided inverse.*

This theorem shows the practical equivalence of the two concepts, deformation retraction and homotopy type. To emphasize this I restate theorem 3.7.

THEOREM 3.8. *Two spaces X and Y belong to the same homotopy type if and only if they can both be imbedded in a third space W in such a way that they are both deformation retracts of W. The dimension of W need not be larger than* max $\{\dim Y, 1 + \dim X\}$.

By induction there follows

(3.9) *If the spaces $X_1, \cdots, X_k$ belong to the same homotopy type then there is a space W of which each X_i is a deformation retract. The dimension of W need not be larger than* $1 + \max_{i=1,\cdots,k} \{\dim X_i\}$, *or than* $\max_{i=1,\cdots,k} \{\dim X_i\}$ *if only one X_i has the maximum dimensionality.*

I conclude this section with an example illustrating the utility of theorem 3.7. Let T be a 2-simplex with vertices a, b, c and let E be the 2-dimensional complex resulting from T by an identification (denoted by i) of the side ab with the side bc and with the side ac. From general theorems it is known that E is contractible. I will now show how to construct, explicitly, a contraction[25] of E. Let A be a small 2-simplex in the interior of $E = i(T)$. The projection f from an interior point of A maps $X = \dot{A}$[26] onto $Y = i(\dot{T})$. Both X and Y are 1-spheres and $Y + C_f = E - (A - \dot{A})$. Since f is homotopic to a homeomorphism of X on Y, f has a 2-sided inverse. Hence, by theorem 3.7, X is a

[24] Equivalently: *Z can be deformed into $C_g - Z$ if and only if g has a right inverse.*

[25] The contractibility of this example was shown by K. Borsuk, *Über das Phänomen der Unzerlegbarkeit in der Polyedertopologie.* Comm. Math. Helv. 8 (1935), §3, p. 143.

[26] The dot denotes the boundary operation, as in Alexandroff and Hopf, loc. cit.

deformation retract of $Y + C_f$. By theorem 1.4 the retracting deformation may be chosen so that it leaves the points of X fixed throughout the deformation. Thus A is a deformation retract of E. Since A is contractible it follows that E is contractible.

4. The Hopf-Pannwitz deformations

A space A is said to be *inessential*[27] *relative to a subset* B if there is a deformation of A into a proper subset of itself such that the points of B remain fixed during the deformation.

THEOREM 4.1. *Let f be a mapping of X into Y such that $Y + C_f$ is an ANR-set and suppose that f has a right inverse. If V is a proper subset of $Y + C_f$* (which contains X) *and h is a mapping of Y into V such that $hf \simeq 1$ in V then $Y + C_f$ is inessential relative to X.*

On choosing Z of theorem 3.1 to be the set V we have that the identity map of X can be extended to a map of $Y + C_f$ into V. Let this extension be denoted by λ. Since f has a right inverse it follows from theorem 3.4 that $Y + C_f$ can be deformed into X. Hence, by 1.1, $\lambda \simeq 1$ in $Y + C_f$. Thus there is a deformation η of $Y + C_f$ such that $\eta_1 = \lambda$. By 1.3, η can be so chosen that $\eta_t \mid X = 1$ for every $0 \leqq t \leqq 1$. Thus $Y + C_f$ is inessential relative to X.

Let X be the graph consisting of two circles S_1 and S_2 joined by an arc and let Y be a 2-cell. Let f be the map of X into Y which maps the arc into a point of the boundary $\dot{Y}$[26] of Y and maps each of the circles homeomorphically onto $\dot{Y}$. Since Y is contractible, f has a right inverse. A map h satisfying the condition of theorem 4.1 is a map of Y into a point of X, where $V = X + Y + C_{f \mid (S_1 + S_2)}$. (It is easy to verify that X can be contracted in V.) Hence, by theorem 4.1, $Y + C_f$ is inessential relative to X. Let K be the space obtained from a torus by spanning a meridian with a 2-cell and an equator with a 2-cell. It is easy to see that the mapping cylinder $Y + C_f$ just constructed is a subset of K in such a way that $(K - (Y + C_f)) \cdot (Y + C_f) = X$. Hence we deduce that K is inessential. This gives a new and simple proof of a deformation theorem of Hopf and Pannwitz.[28]

If Y is a point then a mapping of Y into a point of X is a right inverse of f. Hence, on choosing $h(y) \in X$, we have

THEOREM 4.2. *If f maps the ANR-set X into a point of Y and if X can be contracted in a proper subset of $Y + C_f$ then $Y + C_f$ is inessential relative to X.*

By choosing X to be the graph described above, theorem 4.2 yields a new proof of another example of Hopf and Pannwitz.

If Y is a point and $f(X) \subset Y$ then $Y + C_f = C_f$ is called the *cone* of X. A homogeneous n-dimensional polyhedron K is said to be *closed*[29] if for some coefficient domain there is an n-cycle whose carrier[30] is K.

[27] Alexandroff and Hopf, loc. cit. p. 521.
[28] Ibid, p. 525.
[29] Ibid, p. 274.
[30] Ibid, p. 169 where the carrier for C is denoted by $\mid C \mid$.

THEOREM 4.3. *The homogeneous n-dimensional* $(n > 0)$ *polyhedron* K *is closed if and only if* K *can not be contracted in any proper subset of its cone.*

If K cannot be so contracted that K is closed by theorem 4.2 and a theorem of Hopf and Pannwitz.[31] Suppose K were closed and could be contracted in a proper subset of the cone. Then a continuous $(n + 1)$-chain, covering a proper subset of the cone, could be found whose boundary cycle has K for its carrier. Since every n-simplex of K lies on exactly one $(n + 1)$-simplex of the cone and conversely, this is impossible.

The *absolute boundary*[32] of a homogeneous n-dimensional polyhedron K is made up of those simplexes which, for every coefficient domain, carry the boundary of every n-chain whose carrier is K.

THEOREM 4.4. *A point* p *of a homogeneous n-dimensional* $(n > 0)$ *polyhedron* K *belongs to the absolute boundary of* K *if and only if the boundary of its star* $st(p)$ *can be contracted in a proper subset of* $st(p)$.[33]

Since $st(p)$ is the cone of the boundary of $st(p)$ this is a consequence[34] of theorem 4.3.

5. Special deformation retracts and homotopy type

Let

$$\rho(\langle x, s\rangle) = f(x), \qquad \langle x, s\rangle \in C_f,$$

$$\rho(y) = y, \qquad y \in Y.$$

I shall say that a retracting deformation ξ of $Y + C_f$ into X is *special* if

$$\xi_t \mid X = 1, \qquad 0 \leqq t \leqq 1 \qquad \text{and}$$

$$\rho(\xi_t(f(x))) = \rho(\xi_1(\langle x, t\rangle)).$$

A two-sided inverse g of f I will call *special* if there are homotopies F and G such that in addition to the usual conditions

$$F_0(x) = x, \qquad F_1(x) = gf(x), \qquad G_0(y) = y, \qquad G_1(y) = fg(y)$$

the condition

$$f(F_t(x)) = G_t(f(x))$$

is satisfied.

THEOREM 5. *In order that* f *have a special 2-sided inverse it is necessary and sufficient that there exist a special retracting deformation of* $Y + C_f$ *into* X.

---

[31] Ibid, p. 521.

[32] Ibid, p. 285.

[33] The star of p is the union of the closed simplexes of K which contain p. The boundary of $st(p)$ is the union of those closed simplexes of K which are contained in $st(p)$ and do not contain p.

[34] Ibid, Satz XIV, p. 285.

Suppose first that ξ is a special retracting deformation of $Y + C_f$ into X. Let

$$g(y) = \xi_1(y), \qquad y \in Y,$$
$$F_t(x) = \xi_1(\langle x, t\rangle), \qquad x \in X, \qquad 0 \leqq t \leqq 1,$$
$$G_t(y) = \rho(\xi_t(y)), \qquad y \in Y, \qquad 0 \leqq t \leqq 1.$$

Clearly $F_0(x) = \xi_1(\langle x, 0\rangle) = \xi_1(x) = x$; $F_1(x) = \xi_1(\langle x, 1\rangle) = \xi_1 f(x) = g(f(x))$; $G_0(y) = \rho(\xi_0(y)) = \rho(y) = y$; $G_1(y) = \rho(\xi_1(y)) = \rho(g(y)) = \rho(\langle g(y), 0\rangle) = f(g(y))$. Also $fF_t(x) = f(\xi_1(\langle x, t\rangle)) = \rho(\xi_1(\langle x, t\rangle)) = \rho(\xi_t(f(x))) = G_t(f(x))$.

Conversely suppose g, F and G are given satisfying $f(F_t(x)) = G_t(f(x))$. A retracting deformation ξ of C_f is defined by

$$\begin{aligned}
\xi_u(\langle x, y\rangle) &= F_{(2u+1)t}(x), && 0 \leqq t \leqq 2u/(2u+1), 0 \leqq u \leqq 1/2,\\
&= \langle F_{2u}(x), (2u+1)t - 2u\rangle, && 2u/(2u+1) \leqq t \leqq 1, 0 \leqq u \leqq 1/2,\\
&= F_{2t}(x), && 0 \leqq t \leqq 1/2, 1/2 \leqq u \leqq 1,\\
&= \langle g(f(x)), (2t-1)(2-2u)\rangle, && 1/2 \leqq t \leqq 1, 1/2 \leqq u \leqq 1.
\end{aligned}$$

It is easy to verify that this definition is consistent. When $0 \leqq u \leqq 1/2$ we have $\xi_u(f(x)) = \xi_u(\langle x, 1\rangle) = \langle F_{2u}(x), 1\rangle = f(F_{2u}(x)) = G_{2u}(f(x))$ and when $1/2 \leqq u \leqq 1$ we have $\xi_u(f(x)) = \langle gf(x), 2 - 2u\rangle$. Hence I may consistently define

$$\begin{aligned}
\xi_u(y) &= G_{2u}(y), && 0 \leqq u \leqq 1/2,\\
&= \langle g(y), 2 - 2u\rangle, && 1/2 \leqq u \leqq 1,
\end{aligned}$$

and thus extend ξ to a retracting deformation of $Y + C_f$ into X. For every $0 \leqq u \leqq 1$ we have $\xi_u(x) = \xi_u(\langle x, 0\rangle) = F_0(x) = x$. Thus $\xi_u \mid X = 1$ for every $0 \leqq u \leqq 1$. Finally I show that $\rho(\xi_t(f(x))) = \rho(\xi_1(\langle x, t\rangle))$. When $0 \leqq t \leqq 1/2$ we have $\rho(\xi_t(f(x))) = \rho(\xi_t(\langle x, 1\rangle)) = \rho(\langle F_{2t}(x), 1\rangle) = (F_{2t}(x)) = \rho(\xi_1(\langle x, t\rangle))$ and when $1/2 \leqq t \leqq 1$ we have $\rho(\xi_t(f(x))) = \rho(\xi_t(\langle x, 1\rangle)) = \rho(\langle g(f(x)), 2 - 2t\rangle) = \rho(\langle g(f(x)), 0\rangle) = \rho(\xi_1(\langle x, t\rangle))$.

6. Mappings with ε-inverses for every ε

A homotopy ξ in a metric space B is called an ε-homotopy[35] if $d(\xi_{t_1}(x), \xi_{t_2}(x)) < \varepsilon$ for every $t_1, t_2 \in [0, 1]$ and $x \in \xi^{-1}(B)$. Accordingly if X and Y are metric spaces and f and g are mappings, of X into Y and Y into X respectively, such that $gf \simeq_\varepsilon 1$ ($\simeq_\varepsilon$ denotes ε-homotopy) then I will call g a *left ε-inverse* of f and f a *right ε-inverse* of g. The property of having a left or right ε-inverse for every $\varepsilon > 0$ is topological.

Assume now that the mapping cylinder $Y + C_f$ is metrizable and has been metrized. This is the case, for instance, if X and Y are compacta.

THEOREM 6.1. *A mapping f of X into Y has a left ε-inverse for every $\varepsilon > 0$ if and only if f is a homeomorphism and for every $\varepsilon > 0$ the identity mapping of $f(X)$ is ε-homotopic in $f(X)$ to a map extendable to Y.*

[35] Ibid, p. 343.

If g^ε is a left ε-inverse of f then $f(x_1) = f(x_2)$ implies that $d(x_1, x_2) \leqq d(x_1, g^\varepsilon(f(x_1))) + d(g^\varepsilon(f(x_2)), x_2) < 2\varepsilon$. Hence if f has left ε-inverses for every $\varepsilon > 0$ then f must be one to one. If $\{x_k\} \subset X$ and $f(x_k) \to f(x_0)$ then $g^\varepsilon(f(x_k)) \to g^\varepsilon(f(x_0))$ for every $\varepsilon > 0$ as $k \to \infty$ and $g^\varepsilon(f(x_k)) \to x_k$ uniformly in k as $\varepsilon \to 0$. Hence $x_k \to x_0$ so that f is a homeomorphism. For every $\varepsilon > 0$ there is a $\delta > 0$ such that $fg^\delta f f^{-1} \simeq_\varepsilon f 1 f^{-1} = 1$ in $f(X)$. Hence the mapping $fg^\delta \mid f(X)$ is extendable to Y and also ε-homotopic to the identity.

It is no loss of generality, in the proof of the converse, to assume that $X \subset Y$ and $f = 1$. The maps g^ε of Y into X such that $g^\varepsilon \mid X \simeq_\varepsilon 1$ in X are easily seen to be left ε-inverses of f.

(6.2) *If X is, in addition, an ANR-set then f has a left ε-inverse if and only if f is a homeomorphism and $f(X)$ is a retract of Y.*

This follows from (6.1) and a theorem[14] of Borsuk-Kuratowski-Dowker.

THEOREM 6.3. *In order that f have a right ε-inverse for every $\varepsilon > 0$ it is necessary and sufficient that Y be ε-deformable*[36] *into $C_f - Y$ for every $\varepsilon > 0$.*

If, for every $\delta > 0$, g^δ is a right δ-inverse and G^δ is a δ-homotopy such that $G_0^\delta(y) = y$ and $G_1^\delta(y) = f(g^\delta(y))$ then, for preassigned ε and sufficiently small δ, the deformation ξ^δ defined by

$$\begin{aligned} \xi_t^\delta(y) &= G_{2t}^\delta(y), \qquad y \in Y, \qquad 0 \leqq t \leqq 1/2, \\ &= \langle g^\delta(y), 1 - (2t-1)\delta \rangle, \qquad y \in Y, \qquad 1/2 \leqq t \leqq 1, \end{aligned}$$

is an ε-deformation of Y into $C_f - Y$.

Let $\nu(\langle x, t \rangle) = x$ for every $\langle x, t \rangle \in C_f - Y$. If, for every $\delta > 0$, ξ^δ is a δ-deformation of Y into $C_f - Y$ then, for preassigned ε and sufficiently small δ, the mapping g of Y into X defined by

$$g(y) = \nu(\xi_1^\delta(y)), \qquad y \in Y,$$

is a right ε-inverse of f.

Note that the condition of theorem 6.3 implies that $\overline{f(X)} = Y$. Hence

(6.4) *If X is compact then f has a 2-sided ε-inverse for every $\varepsilon > 0$ if and only if f is a homeomorphism of X onto Y.*

7. Analysis of n-homotopy type

Mappings f and g of A into B are called *n-homotopic*[37] if for every n-dimensional polyhedron P and mapping ϕ of P into A the mappings $f\phi$ and $g\phi$ are homotopic. The symbol $\simeq^n$ will denote n-homotopy.

THEOREM 7.1. *Let X, Y and Z be topological spaces and let θ be a mapping of X into Z. If f is a mapping of X into Y then there is a mapping g of Y into Z satisfying $gf \simeq^n \theta$ if and only if θ is n-homotopic to a mapping which can be extended to $Y + C_f$.*

If θ^* is a mapping of $Y + C_f$ into Z such that $\theta^* \mid X \simeq^n \theta$ then the mapping $g = \theta^* \mid Y$ satisfies $gf \simeq^n \theta$. In fact $gf = \theta^* \mid \langle X, 1 \rangle \simeq \theta^* \mid \langle X, 0 \rangle = \theta^* \mid X \simeq^n \theta$.

[36] I.e. the deformation is an ϵ-homotopy.

[37] R. H. Fox, *On the Lusternik-Schnirelmann Category*, Annals of Math. 42 (1941), p. 344.

Suppose, conversely, that $gf \simeq^n \theta$. Define

$$\theta^*(\langle x, t\rangle) = g(f(x)), \qquad \langle x, t\rangle \in C_f,$$

$$\theta^*(y) = g(y), \qquad y \in Y.$$

Then $\theta^* X = gf \simeq^n \theta$.

THEOREM 7.2. *Let X, Y, Z and θ be as in theorem* 7.1. *If g is a mapping of Y into Z then there is a mapping f of X into Y satisfying $gf \simeq^n \theta$ if and only if θ is n-homotopic in $Z + C_g$ to a mapping of X into Y.*

The proof can be constructed from that of theorem 3.2 by changing the appropriate homotopies to n-homotopies.

If $gf \simeq^n 1$ the mapping g will be called a left n-homotopy inverse and f will be called a right n-homotopy inverse. If f has a 2-sided n-homotopy inverse X and Y will be said to belong to the same n-homotopy type. If X and Y are compact ANR-sets then g is a left (right) inverse of f if and only if g is a left (right) n-homotopy inverse of f for every $n < 1 + \dim X$ (for every $n < 1 + \dim Y$).[38]

(7.3) *f has a left n-homotopy inverse if and only if the identity mapping of X is n-homotopic to a map which is extendable to $Y + C_f$.*

(7.4) *g has a right n-homotopy inverse if and only if $Z + C_g$ can be n-deformed into Y.*

By the argument of (1.1) it follows from (7.3) and (7.4) that

(7.5) *X and Y belong to the same n-homotopy type if and only if there is a space $W \supset X + Y$, of which Y is a deformation retract, and a mapping h of W into X such that $h \simeq^n 1$ in W and $h \mid X \simeq^n 1$ in X.*

For example a 2-sphere X and a point Y belong to the same 1-homotopy type. Here a space W is a 2-cell of which X is the boundary and Y is the center. Spaces which belong to the same n-homotopy type have isomorphic k-dimensional homotopy groups for $k \leq n$.

UNIVERSITY OF ILLINOIS.

---

[38] Ibid, theorem 13, p. 344.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

ON THE DEFORMATION RETRACTION OF SOME FUNCTION SPACES ASSOCIATED WITH THE RELATIVE HOMOTOPY GROUPS

By Ralph H. Fox

(Received October 1, 1942)

The homotopy group $\pi^{n+1}(B, C) = \pi^{n+1}(B, C, d)$, $(n \geqq 1)$, of an arcwise connected[1] space B relative[2] to an arcwise connected[1] subset C with base point $d \in C$ may be defined as the fundamental group of a certain space[3] $\mathfrak{F}^n(B, C, d)$. The group $\pi^{n+1}(B, C)$ is independent of the base point d in the sense that $\pi^{n+1}(B, C, d_1)$ and $\pi^{n+1}(B, C, d_2)$ are isomorphic. Hurewicz demonstrates this by showing that $\mathfrak{F}^n(B, C, d_1)$ and $\mathfrak{F}^n(B, C, d_2)$ belong to the same homotopy type. According to my generalization[4] of Whitehead's theorem these function spaces are therefore deformation retracts of some containing space W. The containing space W constructed by this method is a subset of the function space $\mathfrak{F}^n(B, C, D)$, where D is the arc, with end-points d_1 and d_2 which appears in Hurewicz' (unpublished) proof. This suggests that $\mathfrak{F}^n(B, C, d_1)$ and $\mathfrak{F}^n(B, C, d_2)$ may be deformation retracts of $\mathfrak{F}^n(B, C, D)$ itself.

It will be shown below that this is indeed the case, at least if B and C are compact ANR-sets. (The reader will note that the reasoning can now be reversed to deduce the independence of $\pi^{n+1}(B, C)$ of the base point.) Furthermore deformation retraction of B or C induces deformation retraction of $\mathfrak{F}^n(B, C, d)$, hence leaves $\pi^{n+1}(B, C)$ unaltered. Before plunging into the proof I generalize the spaces $\mathfrak{F}^n(B, C, D)$ by (a) generalizing the antecedent cells and cell-boundaries to arbitrary topological spaces, and (b) removing the dependence of $\mathfrak{F}^n$ on the obviously irrelevant number three. Thus we might loosely describe the investigation as a study of the relationship between deformation retraction in the image space and deformation retraction of the function space.

---

[1] This slight restriction, which is not required for the definition of the group, simplifies our discussion.

[2] The absolute homotopy group $\pi^{n+1}(B) = \pi^{n+1}(B, d)$, whose definition may be obtained from that of the relative group by identifying B with d, may be discussed analogously.

[3] The set of continuous functions defined on a topological space X with values in a metric space Y is denoted, as usual, by the symbol Y^X. If X is compact or Y is bounded the well known formula

$$d(f, g) = \sup_{x \in A}\{d(f(x), g(x))\}$$

makes Y^X a metric space. If either X or Y is compact then topologically equivalent metrics of Y induce equivalent metrics of Y^X so that the topology of Y^X is then independent of the metrization of the topological space Y.

Let E^n denote the n-cell: $0 \leqq x_i \leqq 1$, $i = 1, 2, \cdots, n$ and $\dot{E}^n$ its boundary: $\prod_{i=1}^n x_i(1 - x_i) = 0$. If D is any subset of C, the symbol $\mathfrak{F}^n(B, C, D)$ denotes the subset of B^{E^n} which consists of those mappings f for which $f(\dot{E}^n) \subset C$ and $f(\dot{E}^n - E^{n-1}) \subset D$.

[4] R. H. Fox, *On Homotopy Type and Deformation Retraction*, this volume, p. 45, theorem 3.8.

The three lemmas are contributions to the theory of fibre spaces[5] and, except for notation, are independent of the rest of the paper. With reference to lemma 3 it should be pointed out that a certain theorem of Borsuk[6], published four years before fibre spaces[5] had been discovered, is, when restated, seen to be a far-reaching result on fibre spaces. Borsuk's fibre theorem reads: If B is a compact ANR-set and A' is closed in A then the operation $f \to f \mid A'$[7], $f \in B^A$ is a fibre mapping of B^A into $B^{A'}$. I have recently discovered a very simple proof of this theorem which will appear elsewhere together with a discussion of its place in fibre space theory.

Let A_λ (λ a non-negative integer) denote a topological space and B_λ a metric space with either A_λ compact or B_λ bounded. Consider a decreasing sequence

$$A_\lambda \supset A_{\lambda-1} \supset \cdots \supset A_0 \supset \cdots \supset A_{-\mu}, \qquad (0 \leqq \mu \leqq \infty)$$

of subsets of A_p and a corresponding sequence

$$B_\lambda \supset B_{\lambda-1} \supset \cdots \supset B_0 \supset \cdots \supset B_{-\mu}$$

of subsets of B_λ. Let B_0' be a subset of B_0 which contains B_{-1} if $\mu \neq 0$. For $0 \leqq i \leqq \lambda$ I shall use the symbol $\mathfrak{F}_i$ to denote the subset of $B_i^{A_i}$ which consists of those mappings f for which

$$f(A_j) \subset B_j \quad \text{for} \quad -\mu \leqq j \leqq i,$$

and the symbol $\mathfrak{F}_i'$ for the subset of $\mathfrak{F}_i$ which consists of those mappings $f \in \mathfrak{F}_i$ which satisfy the additional requirement

$$f(A_0) \subset B_0'.$$

Let h be a deformation of B_0 into B_0' which deforms each of the subsets $B_{-1}, \cdots, B_{-\mu}$ within itself. Thus $h \in B_0^{B_0 \times [0,1]}$ such that

$$(1) \qquad \begin{cases} h_0(y) = y & \text{for } y \in B_0, \\ h_1(B_0) \subset B_0', & \\ h_t(B_{-\nu}) \subset B_{-\nu} & \text{for } 0 \leqq t \leqq 1, \quad -\mu \leqq -\nu \leqq 0. \end{cases}$$

The existence of this deformation is assumed from now on.

Theorem 1. *If B_0 is closed and $B_1, \cdots, B_\lambda$ are ANR-sets[8] closed in B_λ then $\mathfrak{F}_\lambda$ can be deformed into $\mathfrak{F}_\lambda'$.*

---

[5] W. Hurewicz and N. E. Steenrod, *Homotopy Relations in Fibre Spaces*, Proc. Nat. Acad. 27 (1941), 61–64. The earlier definitions of H. Whitney (sphere-spaces, sphere-bundles and fibre bundles) required the fibres to belong to the same topological type. Other definitions—see B. Eckmann, *Zur Homotopietheorie gefaserter Räume*, Comm. Math. Helv. 14 (1941), 141–192, for references—require compactness assumptions.

[6] K. Borsuk, *Sur les prolongements des transformations continues*, Fund. Math. 28 (1937), 99–110.

[7] R. H. Fox, *On Homotopy Type and Deformation Retraction*. loc. cit., footnote 10.

[8] R. H. Fox, *A Characterization of Absolute Neighborhood Retracts*, Bull. Am. Math. Soc. 48 (1942), 271–275.

Since B_0 is closed in the ANR-set B_1 the deformation h of B_0 can be extended to a deformation of B_1 (in itself)[9]. Since B_i is closed in the ANR-set B_{i+1} $(i = 1, 2, \cdots, \lambda - 1)$ the same argument shows that when h has been extended stepwise to a deformation of B_i it can be further extended to a deformation of B_{i+1}. Let h^* denote the deformation of B_λ which is the final result of this sequence of extensions. Thus $h^* \epsilon B_\lambda^{B_\lambda \times [0,1]}$ such that

$$(2) \qquad \begin{cases} h_0^*(y) = y \quad \text{for} \quad y \epsilon B_\lambda \\ h_1^*(B_0) \subset B_0' \\ h_t^*(B_i) \subset B_i \quad \text{for} \quad 0 \leqq t \leqq 1, \quad -\mu \leqq i \leqq \lambda. \end{cases}$$

For every $x \epsilon A_0$ and $f \epsilon \mathfrak{F}_\lambda$ define

$$(3) \qquad \phi_t^f(x) = h_t^*(f(x)).$$

The function ϕ is a deformation of $\mathfrak{F}_\lambda$ into $\mathfrak{F}_\lambda'$.

THEOREM 2. *If $A_0, \cdots, A_\lambda$ are ANR-sets closed in A_λ then $\mathfrak{F}_\lambda$ is deformable into $\mathfrak{F}_\lambda'$.*

First I construct a deformation r of A_λ which is a neighborhood retracting deformation of A_0 and which deforms each $A_i(-\mu \leqq i \leqq \lambda)$ within itself. The construction is inductive; the mapping

$${}^0R_t(x) = x \quad \text{for} \quad x \epsilon A_0$$

is a deformation of A_0 which is a neighborhood retracting deformation of A_0 and which deforms each $A_i(-\mu \leqq i \leqq 0)$ within itself. Since A_0 is a neighborhood retract of A_λ there is a closed neighborhood U_0 of A_0 in A_λ and a retraction ρ of U_0 onto A_0. Suppose that ${}^jR \epsilon A_j^{A_j \times [0,1]}$ $(0 \leqq j < \lambda)$ such that

$$\begin{aligned} {}^jR_0(x) &= x \quad \text{for} \quad x \epsilon A_j, \\ {}^jR_t(A_i) &\subset A_i \quad \text{for} \quad 0 \leqq t \leqq 1 \quad \text{and} \quad -\mu \leqq i \leqq j, \\ {}^jR_1(x) &= \rho(x) \quad \text{for} \quad x \epsilon U_j \cdot A_j, \end{aligned}$$

where U_j is a closed neighborhood of A_0 in U_0 (hence in A_λ). Define

$$\begin{aligned} {}^jS_0(x) &= x \quad \text{for} \quad x \epsilon A_{j+1}, \\ {}^jS_t(x) &= {}^jR_t(x) \quad \text{for} \quad (x, t) \epsilon A_j \times [0, 1], \\ {}^jS_1(x) &= \rho(x) \quad \text{for} \quad x \epsilon U_j \cdot A_{j+1}. \end{aligned}$$

Since $A_{j+1} \times [0] + A_j \times [0, 1] + U_j \cdot A_{j+1} \times [1]$ is closed in $A_{j+1} \times [0, 1]$ and A_{j+1} is an ANR-set, ${}^jS$ can be[10] extended to a map ${}^jS^*$ of a neighborhood V_j (into A_{j+1}). Let U_{j+1} be a closed neighborhood of A_0 in U_j (hence in A_λ) such

[9] R. H. Fox, *On Homotopy Type and Deformation Retraction.* loc. cit. footnote 14.

[10] R. H. Fox, *A Characterization of Absolute Neighborhood Retracts,* loc. cit. p. 273.

that

$$U_{j+1} \cdot A_{j+1} \times [0, 1] \subset V_j ,$$

and define

$$^jT_0(x) = x \quad \text{for} \quad x \in A_{j+1} ,$$
$$^jT_t(x) = {}^jS_t^*(x) \quad \text{for} \quad (x, t) \in U_{j+1} \cdot A_{j+1} \times [0, 1].$$

Since ^jT is homotopic to the identity mapping of the closed set $A_{j+1} \times [0] + U_{j+1} \cdot A_{j+1} \times [0, 1]$ and $A_{j+1} \times [0, 1]$ is an ANR it follows[9] that ^jT can be extended to $A_{j+1} \times [0, 1]$. Let ^{j+1}R denote the extended mapping. But $^{j+1}R \in A_{j+1}^{A_{j+1} \times [0,1]}$ such that

$$^{j+1}R_0(x) = x \quad \text{for} \quad x \in A_{j+1} ,$$
$$^{j+1}R_t(A_i) \subset A_i \quad \text{for} \quad 0 \leqq t \leqq 1 \quad \text{and} \quad -\mu \leqq i \leqq j + 1,$$
$$^{j+1}R_1(x) = \rho(x) \quad \text{for} \quad x \in U_{j+1} \cdot A_{j+1} .$$

This completes the induction; let $r = {}^\lambda R$ so that

$$(4) \qquad \begin{cases} r_0(x) = x & \text{for} \quad x \in A_\lambda , \\ r_t(A_i) \subset A_i & \text{for} \quad 0 \leqq t \leqq 1 \quad \text{and} \quad -\mu \leqq i \leqq \lambda, \\ r_1(x) = \rho(x) & \text{for} \quad x \in U_\lambda . \end{cases}$$

Let $\delta \in [0, 1]^{A_\lambda}$ such that $\delta(\overline{A_\lambda - U_\lambda}) = 0$ and $\delta(A_0) = 1$[11]. A deformation of $\mathfrak{F}_\lambda$ into $\mathfrak{F}_\lambda'$ is defined by the formulae

$$(5) \qquad \begin{aligned} \phi_t^f(x) &= f(r_{2t}(x)) \quad \text{for} \quad 0 \leqq t \leqq 1/2, x \in A_\lambda , \\ &= h_{(2t-1)\delta(x)}(f(r_1(x))) \quad \text{for} \quad 1/2 \leqq t \leqq 1, x \in U_\lambda , \\ &= f(r_1(x)) \quad \text{for} \quad 1/2 \leqq t \leqq 1, x \in A_\lambda - U_\lambda , \end{aligned}$$

where $f \in \mathfrak{F}_\lambda$.

Given a topological space X and metric space Y I shall say that a mapping $\pi \in Y^X$ is a *fibre mapping* if X is a fibre space[5] over $\pi(X) \subset Y$ relative to π. This implies, of course, that $\pi(X)$ is open and closed in Y.[5]

LEMMA 1. *If π is a fibre mapping of X into Y and X' is the complete inverse image of some subset Y' of Y then $\pi \mid X'$[7] is a fibre mapping of X' into Y.*

The proof is immediate.

LEMMA 2. *If π_1 is a fibre mapping of X into Y and π_2 is a fibre mapping of Y into Z whose slicing function is uniformly continuous in y and z together then $\pi_2\pi_1 \in Z^X$ is a fibre mapping.*

According to the definition there exist ϵ_1, $\epsilon_2 > 0$ and slicing functions ϕ_1, ϕ_2 such that

$$\phi_1(x, y) \in \pi_1^{-1}(y) \text{ and is defined whenever } d(\pi_1(x), y) < \epsilon_1 ,$$
$$\phi_2(y, z) \in \pi_2^{-1}(z) \text{ and is defined whenever } d(\pi_2(y), z) < \epsilon_2 ,$$
$$\phi_1(x, \pi_1(x)) = x \quad \text{and} \quad \phi_2(y, \pi_2(y)) = y.$$

[11] Urysohn's Lemma.

Choose $\epsilon < \epsilon_2$ so small that $d(\phi_2(x, z), \phi_2(x, z')) < \epsilon_1$ whenever $d(z, z') < \epsilon$; this is possible because of the uniform continuity of ϕ_2. Let

$$\phi(x, z) = \phi_1(x, \phi_2(\pi_1(x), z)). \tag{6}$$

This is defined whenever $d(\pi_2\pi_1(x), z) < \epsilon$. Furthermore

$$\begin{aligned} \pi_2\pi_1\phi(x, z) &= \pi_2(\pi_1\phi_1(x, \phi_2(\pi_1(x), z))) \\ &= \pi_2(\phi_2(\pi_1(x), z)) \\ &= z, \end{aligned}$$

and

$$\begin{aligned} \phi(x, \pi_2\pi_1(x)) &= \phi_1(x, \phi_2(\pi_1(x), \pi_2\pi_1(x))) \\ &= \phi_1(x, \pi_1(x)) \\ &= x. \end{aligned}$$

Let π_{ij} be the mapping of $\mathfrak{F}_i$ into $\mathfrak{F}_j (i \geqq j)$ defined by $\pi_{ij}(f_i) = f_i \mid A_j$[7], $f_i \in \mathfrak{F}_i$.

LEMMA 3. *If $A_0, \cdots, A_{\lambda-1}$ are closed and $B_1, \cdots, B_\lambda$ are compact ANR-sets then $\pi_{\lambda 0}$ is a fibre mapping*[12].

By the proof of a theorem of Borsuk[6] [in particular, formulae (8) p. 101 (which should read: $\varphi^*(p) = 1 - r_{01}[\bar{\varphi}(p) + \overset{*}{\varphi_0}(p) - \bar{\varphi}_0(p)]$), l. 11 p. 102 (which should read $\bar{\varphi}(x) = \{\varphi^{(*,n)}(x)\}$) and l. 6. p. 103] the mapping $f_i \rightarrow f_i \mid A_{i-1}$[7], $f_i \in B_i^{A_i}$, $1 \leqq i \leqq \lambda$ is a fibre mapping. Moreover [the previously mentioned formulae, (1) p. 100 and modification of the proof of theorem 2 to the extent of replacing the neighborhood U last line p. 102 by a closed neighborhood so that the retraction r is uniformly continuous] the slicing function for this fibre mapping is uniformly continuous.

But the inverse image of $\mathfrak{F}_{i-1}$ under the mapping is precisely $\mathfrak{F}_i$. Hence, by lemma 1, $\pi_{i\,i-1}$ is a fibre mapping. Hence, by lemma 2, $\pi_{\lambda 0} = \pi_{10}\pi_{21} \cdots \pi_{\lambda\lambda-1}$ is a fibre mapping.

THEOREM 3. *If $A_0, \cdots, A_{\lambda-1}$ and B_0 are closed and $B_1, \cdots, B_\lambda$ are compact ANR-sets and if $h_t(y) = y$ for every $y \in B_0'$, $0 \leqq t \leqq 1$ then $\mathfrak{F}_\lambda'$ is a deformation retract of $\mathfrak{F}_\lambda$.*

Let $\psi \in \mathfrak{F}_0^{\mathfrak{F}_\lambda \times [0,1]}$ defined by

$$\psi_t^f(x) = h_t f(x) \quad \text{for} \quad 0 \leqq t \leqq 1, \quad x \in A_0 \quad \text{and} \quad f \in \mathfrak{F}_\lambda. \tag{7}$$

I shall show that $\psi(\mathfrak{F}_\lambda \times [0, 1]) \subset \pi_{\lambda 0}(\mathfrak{F}_\lambda)$. In fact

$$\begin{aligned} \psi_t^f \quad &= \pi_{00}(\psi_t^f), \\ \psi_0^f(x) &= f(x) \quad \text{for} \quad x \in A_0, \end{aligned}$$

---

[12] Since the image set of a fibre mapping is open and closed it follows from R. H. Fox, *A Characterization of Absolute Neighborhood Retracts*, loc. cit. footnote 3 that the compactness of B_λ is essential to this lemma.

and $\psi_t^f \in \mathfrak{F}_0$ for every fixed $t \in [0, 1]$. Suppose, inductively, that for every fixed $f \in \mathfrak{F}_\lambda$ there is a map ${}^i\psi = {}^i\psi^f \in B_i^{A_i \times [0,1]}$, $0 \leqq i < \lambda$, such that

$$\pi_{i0}({}^i\psi_t^f) = {}^0\psi_t^f = \psi_t^f \qquad \text{for } 0 \leqq t \leqq 1,$$

$${}^i\psi_0^f(x) = f(x) \qquad \text{for } x \in A_i,$$

$${}^i\psi_t^f \in \mathfrak{F}_i \qquad \text{for every fixed } t \in [0, 1].$$

Let

$${}^i\xi_t^f(x) = \begin{cases} {}^i\psi_t^f(x) & \text{for } (x, t) \in A_i \times [0, 1], \\ f(x) & \text{for } (x, t) \in A_{i+1} \times [0]. \end{cases}$$

Since ${}^i\xi^f$ is homotopic to the map ${}^i\eta^f$ defined by

$${}^i\eta_t^f(x) = f(x) \quad \text{for} \quad (x, t) \in A_{i+1} \times [0] + A_i \times [0, 1],$$

since ${}^i\eta^f$ can be extended to $A_{i+1} \times [0, 1]$ and since ${}^i\xi^f$ maps the closed subset $A_{i+1} \times [0] + A_i \times [0, 1]$ of $A_{i+1} \times [0, 1]$ into the compact *ANR*-set B_{i+1}, it follows[9] that ${}^i\xi^f$ can be extended to a map ${}^{i+1}\psi^f$ of $A_{i+1} \times [0, 1]$ into B_{i+1}. But $\pi_{i+1\,0}({}^{i+1}\psi_t^f) = \pi_{i0}(\pi_{i+1\,i}({}^{i+1}\psi_t^f)) = \psi_t^f$ for $0 \leqq t \leqq 1$,

$${}^{i+1}\psi_0^f(x) = {}^i\xi_0^f(x) = f(x) \quad \text{for} \quad x \in A_{i+1},$$

$${}^{i+1}\psi_t^f \in \mathfrak{F}_{i+1} \quad \text{for every} \quad t \in [0, 1],$$

and this completes the induction.

Since B_0 is compact, $\psi \in \mathfrak{F}_0^{\mathfrak{F}_\lambda \times [0,1]}$ is a uniform homotopy[5]. By the previous lemma $\pi_{\lambda 0}$ is a fibre mapping. Therefore, since $\psi(\mathfrak{F}_\lambda \times [0, 1]) \subset \pi_{\lambda 0}(\mathfrak{F}_\lambda)$, it follows by the covering homotopy theorem[5] that there is a mapping $\phi \in \mathfrak{F}_\lambda^{\mathfrak{F}_\lambda \times [0,1]}$ such that

$$\phi_0^f = f \quad \text{for every} \quad f \in \mathfrak{F}_\lambda,$$

$$\pi_{\lambda 0}(\phi) = \psi,$$

$$\phi_1^f \in \mathfrak{F}_\lambda' \quad \text{for every} \quad f \in \mathfrak{F}_\lambda.$$

Thus ϕ is a deformation of $\mathfrak{F}_\lambda$ (within itself) into $\mathfrak{F}_\lambda'$.

If $f \in \mathfrak{F}_\lambda'$ then $f(x) \in B_0'$ for every $x \in A_0$, hence $\psi_t^f = h_t f$ is independent of t. Hence, by the covering homotopy theorem[5], ϕ_t^f is independent of t. Hence, for every $f \in \mathfrak{F}_\lambda'$,

$$\phi_t^f = \phi_0^f = f.$$

Hence ϕ is a retracting deformation (which leaves the points of $\mathfrak{F}_\lambda'$ invariant).

COROLLARY. *$\mathfrak{F}_\lambda'$ is a deformation retract of $\mathfrak{F}_\lambda$ if $A_0, \cdots, A_{\lambda-1}$ are closed, $B_0, B_0', B_1, \cdots, B_\lambda$ are compact ANR-sets and h is a retracting deformation.*

For under these conditions h may be assumed to satisfy[13] the condition of the preceding theorem in addition to (1).

UNIVERSITY OF ILLINOIS, URBANA, ILL.

[13] Ibid, Theorem 1.4.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

ON PERMUTATION GROUPS OF PRIME DEGREE AND RELATED CLASSES OF GROUPS

By Richard Brauer*

(Received June 17, 1942)

Introduction

The transitive permutation groups of prime degree p appear as the Galois groups of the irreducible algebraic equations $f(x) = 0$ of degree p. This is the reason that these groups have been the subject of a large number of investigations.[1] However, only few results of a general nature have been obtained. In the present paper, the theory of group representations[2] will be applied in order to derive some new theorems concerning the structure of these groups. Actually, the method can be used for the study of a wider class of groups, viz. the groups $\mathfrak{G}$ of finite order g which have the following property:

(*) *The group $\mathfrak{G}$ contains elements P of prime order p which commute only with their own powers P^i.*

It is clear that transitive permutation groups of degree p have the property (*). Secondly, the doubly transitive permutation groups of degree $p - 1$ are of this type.[3] A third example is furnished by the irreducible linear groups in a p-dimensional vector space whose center consists of the unit element only, in particular by the simple linear irreducible groups in p dimensions (cf. section 7).

It is easily seen (section 1) that the order g of a group $\mathfrak{G}$ with the property (*) is of the form

$$g = (p - 1)p(1 + np)/t \tag{1}$$

where t and n are integers and where t divides $p - 1$. The group $\mathfrak{G}$ contains exactly $1 + np$ conjugate subgroups of order p, and each of them has a normalizer of order $p(p - 1)/t$. In section 2, the normal subgroups of $\mathfrak{G}$ are studied, in particular the first commutator-subgroup $\mathfrak{G}'$ and the second commutator-subgroup $\mathfrak{G}''$ of $\mathfrak{G}$. Two cases must be distinguished:

Case I. *The group $\mathfrak{G}$ contains a normal subgroup $\mathfrak{S}$ of order* $1 + np$.

We shall show that $\mathfrak{G}/\mathfrak{S}$ then is a metacyclic group of order $p(p - 1)/t$; the group $\mathfrak{S}$ possesses an outer automorphism of order p which leaves only the

* Fellow of the John Simon Guggenheim Memorial Foundation.

[1] We may mention here the work of Mathieu, C. Jordan, Sylow, Frobenius, Burnside, G. A. Miller. Cf. also E. Pascal, Repertorium der höheren Mathematik, Vol. I, part 1, 2nd German edition, Leipzig 1910.

[2] In this paper, the notation "representation of a group" always means a representation of the group by linear transformations of a vector space over the field of complex numbers (or an algebraically closed field of characteristic 0). By a "vector-space" we always mean a vector-space over this field.

[3] For this class of groups, cf. G. Frobenius, Sitzungsberichte der Preussischen Akademie, Berlin 1902, p. 351.

unit element fixed. For $t < p - 1$, we have $\mathfrak{S} = \mathfrak{G}''$, and $\mathfrak{G}'$ has the order $p(1 + np)$. For $t = p - 1$, we have $\mathfrak{S} = \mathfrak{G}'$. Unless n is of the form

$$n = u + m + ump \qquad (u, m \text{ positive integers}), \tag{2}$$

$\mathfrak{S}$ is a minimal normal subgroup of $\mathfrak{G}$ (for $n \neq 0$).

This case I is of relatively small interest. In particular, when $\mathfrak{G}$ is a transitive permutation group of degree p, $\mathfrak{S}$ consists in this case I only of the unit element 1. If $\mathfrak{G}$ is a doubly transitive group of degree $p + 1$ or an irreducible linear group in a p-dimensional space with center 1, then $\mathfrak{S}$ must be abelian.

CASE II. *The group $\mathfrak{G}$ does not contain a normal subgroup of order* $1 + np$.

Here, we shall have $\mathfrak{G}' = \mathfrak{G}''$. The group $\mathfrak{G}'$ itself satisfies the condition (*); its order g' is of the form

$$g' = (p - 1)p(1 + np)/t' \tag{3}$$

where n is the same number as in (1). The number t' divides $p - 1$ and is divisible by t; we have $t' \neq p - 1$. If n is not of the form (2), in particular, if $n < p + 2$, then $\mathfrak{G}'$ is simple.

In the later sections, we shall assume that $\mathfrak{G}$, besides condition (*), satisfies the following condition

(**) *The commutator-subgroup $\mathfrak{G}'$ of $\mathfrak{G}$ is equal to $\mathfrak{G}$.*

By this condition (**), groups $\mathfrak{G}$ for which we have case I are excluded. If we have case II, the group $\mathfrak{G}'$ satisfies both conditions (*) and (**), and our theory can be applied to $\mathfrak{G}'$. From $\mathfrak{G}'$, the group $\mathfrak{G}$ can be obtained by a cyclic extension; the value of n remains unchanged.

Our main result (section 5), is: If a group $\mathfrak{G}$ satisfies the conditions (*) and (**), and if $n \geqq (p + 3)/2$, then n can be represented by the following rational function $F(p, u, h)$

$$n = F(p, u, h) = \frac{puh + u^2 + u + h}{u + 1} \tag{4}$$

where u and h are positive integers, and where $u + 1$ divides $h(p - 1)$. If $\mathfrak{G}$ satisfies the conditions (*) and (**), and if $n < (p + 3)/2$, we must have one of the following two cases:

(a) $n = 1, \quad t = 2, \quad \mathfrak{G} = LF(2, p), \qquad (p > 3).$

(b) $n = (p - 3)/2, \quad t = (p - 1)/2, \quad \mathfrak{G} = LF(2, 2^\mu)$ where $p = 2^\mu + 1$

is a Fermat prime, $p > 3$.[4]

In a later paper, the values n with $(p + 3)/2 \leqq n \leqq p + 2$ will be discussed.

It had been shown by Frobenius that $LF(2, p)$ is the only simple group of order $p(p - 1)(p + 1)/2$. In section 6 we drop the assumption (*) and prove that the groups $\mathfrak{G} = LF(2, p)$ and $LF(2, 2^\mu)$. $(2^\mu + 1 = p)$ are the only simple

[4] That permutation groups of degree p with the value $n = (p - 3)/2$ exist for these primes p, was mentioned by Frobenius, loc. cit.

groups of an order $p(p-1)(1+mp)/\tau$ with $m < (p+3)/2$, (p a prime, τ, m not-negative integers, $\tau \mid (p-1)$); if for a simple group of this order we have $m \geqq (p+3)/2$, then m must be of the form $m = F(p, u, h)$ where u and h are positive integers.

1. Preliminary remarks

Let $\mathfrak{G}$ be a group of finite order g which satisfies the condition (*), i.e. which contains elements P of prime order p whose centralizer consists of the powers of P only. If $\mathfrak{P}$ is a p-Sylow subgroup of $\mathfrak{G}$ which contains P, then the order of $\mathfrak{P}$ cannot be larger than p, since otherwise the order of the centralizer of P in $\mathfrak{P}$ would be larger than p. Hence $g \not\equiv 0 \pmod{p^2}$, $\mathfrak{P} = \{P\}$. The number of subgroups conjugate to $\mathfrak{P}$ is of the form $1 + np$ where n is a non-negative integer. The order of the normalizer $\mathfrak{N} = \mathfrak{N}(\mathfrak{P})$ of $\mathfrak{P}$ then is $g/(1+np)$. But since $\mathfrak{P}$ is a cyclic group of order p, and since $\mathfrak{N}$ also satisfies the condition (*), we readily see that $\mathfrak{N}$ can be generated by $\mathfrak{P}$ and another element Q such that

$$P^p = 1, \qquad Q^q = 1, \qquad Q^{-1}PQ = P^{\gamma^t} \tag{5}$$

where γ is a primitive root (mod p), and where t and q are positive integers such that

$$tq = p - 1. \tag{6}$$

The group $\mathfrak{G}$ then contains exactly t classes of conjugate elements of order p. For the order of $\mathfrak{G}$, we obtain

$$g = (p-1)p(1+np)/t = qp(1+np). \tag{7}$$

Hence we have

THEOREM 1. *If $\mathfrak{G}$ is a group of finite order g which contains an element P of prime order p which commutes only with its own powers (condition (*)), then $g = (p-1)p(1+np)/t$, where n and t are integers, and t divides $p - 1$. The group $\mathfrak{G}$ contains exactly $1 + np$ subgroups of order p, and t is the number of classes of conjugate elements of order p in $\mathfrak{G}$.*

Since g contains the prime p only to the first power, the results of an earlier paper[5] can be applied. For the sake of convenience we mention those facts which will be needed.

The ordinary irreducible representations of $\mathfrak{G}$ are of four different types:

(I) Representations $\mathfrak{A}_\rho$ of a degree $a_\rho = u_\rho p + 1 \equiv 1 \pmod{p}$. Denote by $A_\rho(G)$ the value of the character A_ρ of $\mathfrak{A}_\rho$ for an element G of $\mathfrak{G}$. Then

$$A_\rho(P^i) = 1 \qquad \text{(for } i \not\equiv 0 \pmod{p}\text{)}. \tag{8, I}$$

II. Representations $\mathfrak{B}_\sigma$ of a degree $b_\sigma = v_\sigma p - 1 \equiv -1 \pmod{p}$. If $B_\sigma(G)$ is the character of $\mathfrak{B}_\sigma$, we have

[5] R. Brauer, *On groups whose order contains a prime number to the first power*, American Journal of Mathematics vol. 54 (1942) part I p. 401, part II, p. 421. I refer to these papers as [1] and [2].

(8, II) $$B_\sigma(P^i) = -1 \qquad (\text{for } i \not\equiv 0 \pmod p).$$

(III) Representations $\mathfrak{C}$ of a degree c which is not congruent to 0, 1, $-1 \pmod p$ for $t \neq 1$.[6] There exist exactly t such representations $\mathfrak{C}, \mathfrak{C}', \cdots, \mathfrak{C}^{(t-1)}$, and they are algebraically conjugate. The degree c is of the form

$$c = (wp + \delta)/t, \qquad \delta = \pm 1$$

where w is a positive integer. If ϵ is a primitive pth root of unity, suitably chosen, we have for the character $C(G)$ of $\mathfrak{C}$:

(8, III) $$C(P^i) = (-\delta) \sum_{\mu=0}^{q-1} \epsilon^{i\gamma^{\mu t}} \qquad \text{for } i \not\equiv 0 \pmod p.$$

We denote the expression on the right side by $(-\delta)\eta_i$ so that η_i is a Gaussian period of length $q = (p-1)/t$.

(IV) Representation $\mathfrak{D}_\tau$ of a degree $d_\tau = px_\tau \equiv 0 \pmod p$. If $D_\tau(G)$ is the character of $\mathfrak{D}_\tau$, then

(8, IV) $$D_\tau(P^i) = 0 \qquad \text{for } i \not\equiv 0 \pmod p.$$

If we have α representation $\mathfrak{A}_\rho$, $\rho = 1, 2, \cdots, \alpha$, and β representations $\mathfrak{B}_\sigma$, $\sigma = 1, 2, \cdots, \beta$, we have

(9) $$\alpha + \beta = q = (p-1)/t.$$

Furthermore, for elements G of an order prime to p, we have

(10) $$\sum_{\rho=1}^{\alpha} A_\rho(G) + \delta C^{(\nu)}(G) = \sum_{\sigma=1}^{\beta} B_\sigma(G).$$

In particular, for $G = 1$, this gives

(11) $$\sum_\rho a_\rho + \delta c = \sum_\sigma b_\sigma .$$

It is well known that the degrees a_ρ, b_σ, c, d_τ divide the order g of $\mathfrak{G}$ and that g is equal to the sum of the squares of all the degrees, i.e.

(12) $$\sum_\rho a_\rho^2 + \sum_\sigma b_\sigma^2 + tc^2 + \sum_\tau d_\tau^2 = g.$$

It is often convenient to set (as above)

(13) $$a_\rho = u_\rho p + 1, \quad b_\sigma = v_\sigma p - 1, \quad c = (wp + \delta)/t, \quad d_\tau = x_\tau p, \quad (\delta = \pm 1).$$

On substituting these values in (11) and taking (9) into account, we easily obtain

(14) $$\sum_\rho u_\rho + \frac{\delta w + 1}{t} = \sum_\sigma v_\sigma .$$

[6] In the case $t = 1$, $\mathfrak{C}$ can be chosen arbitrarily among the p irreducible representations of degrees not divisible by p. We then choose $\mathfrak{C}$ so that its degree c is of the form $c \equiv -1 \pmod p$. This is always possible. The results given in (III) remain valid for this $\mathfrak{C}$. We then have $\delta = -1$, and $C(P^i)$ is rational.

Substitute the values (13) in (12) and use (9) and (14). A simple computation gives

$$\sum u_\rho^2 + \sum v_\sigma^2 + \frac{w^2}{t} + \sum x_\tau^2 = \frac{pn - n + 1}{t}. \tag{15}$$

2. Normal subgroups of $\mathfrak{G}$

The number of representations of degree 1 of any group $\mathfrak{G}$ is equal to the index $(\mathfrak{G}:\mathfrak{G}')$ of the commutator-subgroup $\mathfrak{G}'$ of $\mathfrak{G}$. In our case, for $t \neq 1$, $t \neq p - 1$, only the representations $\mathfrak{A}_\rho$ can have degree 1. By (9), their number is at most $(p - 1)/t$. For $t = 1$, we may choose $\mathfrak{C}$ so that its degree is different from 1, and the same argument holds. If $t = p - 1$, then $\alpha + \beta = 1$, cf. (9). Since the 1-representation $G \rightarrow 1$ appears among the $\mathfrak{A}_\rho$, we have $\alpha = 1, \beta = 0$, $a_1 = 1$, and (11) gives $c = 1$. Consequently, $\mathfrak{G}$ has exactly p representations of degree p. We thus proved

THEOREM 2. *If the group $\mathfrak{G}$ satisfies the condition (*), then the index $(\mathfrak{G}:\mathfrak{G}')$ of the commutator subgroup $\mathfrak{G}'$ in $\mathfrak{G}$ satisfies the relation*

$$(\mathfrak{G}:\mathfrak{G}') \leqq (p - 1)/t \qquad \textit{if } t \neq p - 1,$$

$$(\mathfrak{G}:\mathfrak{G}') = p, \qquad \textit{if } t = p - 1.$$

If $\mathfrak{G}$ has a normal subgroup $\mathfrak{S}$, then any representation of the factor group $\mathfrak{G}/\mathfrak{S}$ may be considered as a representation of $\mathfrak{G}$. On account of this remark, we prove easily

THEOREM 3. *Let $\mathfrak{G}$ be a group of order g which satisfies condition (*). If $\mathfrak{S}$ is a normal subgroup of an order s divisible by p, then $\mathfrak{S}$ contains the commutator-subgroup $\mathfrak{G}'$ of $\mathfrak{G}$.*

PROOF: Let $\mathfrak{Z}$ be an irreducible representation of $\mathfrak{G}/\mathfrak{S}$ of degree z; let $\zeta(G)$ be the character of the corresponding representation of $\mathfrak{G}$. Since the element P of order p must belong to $\mathfrak{S}$, we have $\zeta(P) = z$. The formulas (8) then show that $z \geqq 2$ is impossible; every irreducible representation of $\mathfrak{G}/\mathfrak{S}$ is of degree 1. Hence $\mathfrak{G}/\mathfrak{S}$ is abelian, i.e. $\mathfrak{S}$ contains $\mathfrak{G}'$, q.e.d.

We now treat normal subgroups of an order which is relatively prime to p. We have

THEOREM 4. *Let $\mathfrak{G}$ be a group of order g which satisfies condition (*). If $\mathfrak{S}$ is a normal subgroup of an order s which is not divisible by p, then s divides $1 + np$ and we have $s \equiv 1 \pmod p$.*[7] *The group $\mathfrak{G}/\mathfrak{S}$ itself satisfies condition (*), and the number t^* of classes containing conjugate elements of order p is the same as the analogous number for $\mathfrak{G}$; i.e. $t = t^*$. The group $\mathfrak{S}$ is contained in the kernel of the representations $\mathfrak{A}_\rho$, $\mathfrak{B}_\sigma$, $\mathfrak{C}^{(\nu)}$ (§1).*

PROOF: The order of $\mathfrak{G}/\mathfrak{S}$ is divisible by p; we have $t^* > 0$. Obviously, $t^* \leqq t$. Consider now the representations of the first p-block of $\mathfrak{G}/\mathfrak{S}$.[8] If

[7] The numbers n and t are defined in theorem 1.

[8] cf. [1], section 8.

$t^* \neq 1$, we find t^* representations whose characters take on distinct algebraically conjugate values for an element of order p. These representations yield t^* representations of $\mathfrak{G}$ with the same property, and the formulas (8) show that $t = t^*$. If $t^* = 1$, we find p representations of $\mathfrak{G}/\mathfrak{S}$ whose characters have non-vanishing rational values for an element of order p. Since this again gives p representations of $\mathfrak{G}$ with the corresponding property, we must have $t = 1$. This shows that $t = t^*$ in any case. The first p-block of $\mathfrak{G}/\mathfrak{S}$ now accounts for $t + (p - 1)/t$ representations of $\mathfrak{G}$ of a degree prime to p. But this is the full number of such representations (cf. §1), and hence $\mathfrak{G}/\mathfrak{S}$ does not contain any other p-block of lowest kind. Then the order of the centralizer of a p-Sylow group of $\mathfrak{G}/\mathfrak{S}$ is equal to p;[9] i.e. $\mathfrak{G}/\mathfrak{S}$ satisfies the condition (*). At the same time we proved that $\mathfrak{S}$ is contained in the kernel of all representations A_ρ, B_σ, $C^{(\nu)}$.

The order g/s of $\mathfrak{G}/\mathfrak{S}$ can be written in the form

$$g/s = (p - 1)p(1 + mp)/t \tag{16}$$

where m is a non-negative integer. Comparison of (16) with (7) shows that $s = (1 + np)/(1 + mp)$. Hence s divides $1 + np$, and we have $s \equiv 1 \pmod p$. This proves theorem 4.

Corollary 1. *Any normal subgroup $\mathfrak{S}$ of $\mathfrak{G}$ of an order prime to p possesses an outer automorphism of order p which leaves only the unit element invariant.*

Proof: Transformation of $\mathfrak{S}$ with an element P of order p in $\mathfrak{G}$ defines such an automorphism.—This shows again that $s \equiv 1 \pmod p$.

Corollary 2. *The kernel of any representation $\mathfrak{B}_\sigma$, $\mathfrak{C}^{(\nu)}$ is the (unique) maximal normal subgroup $\mathfrak{S}^*$ of an order prime to p. The same holds for the kernel of $\mathfrak{A}_\rho$, if the degree a_ρ is not* 1.

Proof: As shown above, $\mathfrak{S}^*$ will belong to each such kernel. But the formulas (8, I), (8, II), (8, III) show that the kernel cannot contain elements of order p, i.e. the kernel itself has an order prime to p, and it coincides therefore with $\mathfrak{S}^*$.

Corollary 3. *We have $\mathfrak{S}^* \subseteq \mathfrak{G}'$. If $t \neq p - 1$, the group $\mathfrak{G}'/\mathfrak{S}^*$ is simple. If $t = p - 1$, $\mathfrak{G}' = \mathfrak{S}^*$.*

Proof: The group $\mathfrak{G}'$ can be defined as the intersection of the kernels of the representations of degree 1. Theorem 4 then gives $\mathfrak{S}^* \subseteq \mathfrak{G}'$. If $t \neq p - 1$, the order of $\mathfrak{G}'$ is divisible by p (cf. theorem 2). From theorem 3 and the definition of $\mathfrak{S}^*$ it follows that no normal subgroup of $\mathfrak{G}$ lies between $\mathfrak{G}'$ and $\mathfrak{S}^*$. Then the group $\mathfrak{G}'/\mathfrak{S}^*$ is a minimal normal subgroup of $\mathfrak{G}/\mathfrak{S}^*$, and hence $\mathfrak{G}'/\mathfrak{S}^*$ is a direct product of isomorphic simple groups. But since the order of $\mathfrak{G}'/\mathfrak{S}^*$ contains p to the first power, this implies that $\mathfrak{G}'/\mathfrak{S}^*$ is simple. If $t = p - 1$, theorem 2 shows that $\mathfrak{G}' = \mathfrak{S}^*$.

Corollary 4. *If n is not of the form $n = u + m + ump$ (u, m positive integers), in particular if $n < p + 2$, then $\mathfrak{G}$ does not contain a normal subgroup $\mathfrak{S} \neq \{1\}$ of an order s smaller than $1 + np$.*

[9] cf. [1], theorem 1.

PROOF: If $s \equiv 0 \pmod{p}$, theorems 2 and 3 give $s \geqq (g:(p-1)/t) = p(1+np)$. If $s \not\equiv 0$, theorem 4 shows that s is of the form $1 + up$ where u is a positive integer. From (7) and (16) we obtain

$$1 + np = (1 + mp)(1 + up).$$

Hence $n = u + m + ump$. Under our present assumption, we must have $m = 0$, i.e. $s = 1 + np$ and this proves the corollary.

We now distinguish two cases:

CASE I. *The Group $\mathfrak{G}$ contains a normal subgroup of order $1 + np$.*

CASE II. *The group $\mathfrak{G}$ does not contain a normal subgroup of order $1 + np$.* In other words, in case I the order s^* of $\mathfrak{S}^*$ is equal to $1 + np$ while in case II s^* is smaller than $1 + np$.

THEOREM 5. *We have case I, if and only if one of the following two sets of conditions holds*

(a) $$t = p - 1.$$

(b) $t < p - 1$, *and the first and second commutator subgroups $\mathfrak{G}'$ and $\mathfrak{G}''$ of $\mathfrak{G}$ are different.*

In case (a), $\mathfrak{G}'$ has the order $1 + np$, and $\mathfrak{G}/\mathfrak{G}'$ is cyclic of order p. In case (b), the group $\mathfrak{G}''$ has the order $1 + np$ and $\mathfrak{G}'$ has the order $p(1 + np)$; $\mathfrak{G}/\mathfrak{G}''$ is metacyclic and can be defined by the equations (5).

PROOF: The case $t = p - 1$ is trivial, cf. theorem 2 and (7); we may assume $t < p - 1$. If $\mathfrak{S}^*$ is an invariant subgroup of order $1 + np$ in $\mathfrak{G}$, then $\mathfrak{G}/\mathfrak{S}^*$ is a group of order $p(p-1)/t$, which satisfies condition (*) and in which t classes of conjugate elements contain elements of order p. Hence $\mathfrak{G}/\mathfrak{S}^*$ contains a subgroup of type (5), and since this subgroup has order $p(p-1)/t$, the group $\mathfrak{G}/\mathfrak{S}^*$ itself is a metacyclic group of type (5). In particular, $\mathfrak{G}/\mathfrak{S}^*$ contains a normal subgroup $\mathfrak{G}_1/\mathfrak{S}^*$ of index $(p-1)/t$. Then $\mathfrak{G}_1$ is a normal subgroup of index $(p-1)/t$ of $\mathfrak{G}$, and theorems 2 and 3 now show that $\mathfrak{G}_1 = \mathfrak{G}'$. We may apply theorem 2 to $\mathfrak{G}'$ which again satisfies condition (*). Since $\mathfrak{G}'$ contains a normal subgroup $\mathfrak{S}^*$ of index p, this group $\mathfrak{S}^*$ must be the commutator subgroup $\mathfrak{G}''$ of $\mathfrak{G}'$. Conversely, assume that $\mathfrak{G}' \neq \mathfrak{G}''$. According to theorem 2 we have $(\mathfrak{G}:\mathfrak{G}') \leqq (p-1)/t$. The order of $\mathfrak{G}'$ then is divisible by p, and $\mathfrak{G}'$ also satisfies the condition (*). If the index $(\mathfrak{G}':\mathfrak{G}'')$ was prime to p, the group $\mathfrak{G}''$ would have an order divisible by p, and theorem 3 would give $\mathfrak{G}'' \supseteq \mathfrak{G}'$, i.e. $\mathfrak{G}'' = \mathfrak{G}'$. Hence $(\mathfrak{G}':\mathfrak{G}'')$ is divisible by p. Now theorem 2, applied to $\mathfrak{G}'$, gives $(\mathfrak{G}':\mathfrak{G}'') = p$ and, therefore, $(\mathfrak{G}:\mathfrak{G}'') \leqq p(p-1)/t$. However, theorem 4 shows that the order of the normal subgroup $\mathfrak{G}''$ of $\mathfrak{G}$ must divide $1 + np$. As $\mathfrak{G}$ has the order $p(p-1)(1+np)/t$, we now see that $\mathfrak{G}''$ has the order $1 + np$. This completes the proof of theorem 5.

COROLLARY 5. *In case II, the order g' of the group $\mathfrak{G}'$ is given by*

(17) $$g' = (p-1)p(1+np)/t'$$

where n is the same number as in (7) *and t' denotes the number of classes of conjugate elements in $\mathfrak{G}'$ which contain elements of order p. We have $t \mid t'$, $t' \mid (p - 1)$, and $t \leqq t' < p - 1$. Furthermore, $\mathfrak{G}' = \mathfrak{G}''$. The group $\mathfrak{G}/\mathfrak{G}'$ is cyclic.*

Proof: It follows from theorem 5 that $t < p - 1$ and that $\mathfrak{G}' = \mathfrak{G}''$. The group $\mathfrak{G}'$ also satisfies condition (*), and since it contains all subgroups of order p of $\mathfrak{G}$, we obtain (17). The number t divides t', because g' divides g. If we had $t' = p - 1$, theorem 2 would give $\mathfrak{G}' \neq \mathfrak{G}''$. The element Q in (5) has the property that its (t'/t)th power is the first power which belongs to $\mathfrak{G}'$. This shows that $\mathfrak{G}/\mathfrak{G}'$ is cyclic.

Corollary 6. *If n is not of the form $n = u + m + ump$, (u, m positive integers), in particular if $n < p + 2$, then $\mathfrak{G}'$ is simple in case II.*

Proof: Corollary 4 shows that $\mathfrak{S}^* = \{1\}$, and corollary 3 now proves the statement.

Finally, we treat the three kinds of groups mentioned in the introduction. We prove

Theorem 6. *If $\mathfrak{G}$ is a transitive permutation group of degree p (p a prime number), then $\mathfrak{G}$ does not contain any normal subgroup of an order prime to p and different from $\{1\}$; the group $\mathfrak{G}'$ is simple (or of order 1). If $\mathfrak{G}$ is a doubly transitive group of degree $p + 1$ or if $\mathfrak{G}$ is an irreducible linear group with center $\{1\}$ in a p-dimensional space, then any normal subgroup of an order prime to p is abelian. In all these cases, the composition series of $\mathfrak{G}$ has at most one non-cyclic factor group.*

Proof: If $\mathfrak{G}$ is a transitive permutation group of degree p, then $\mathfrak{G}$ possesses a reducible (1–1)-representation $\mathfrak{Z}$ of degree p, whose character has the value 0 for elements of order p. As shown by the formulas (8, I), this representation cannot consist of representations $\mathfrak{A}_\rho$ exclusively; furthermore it cannot contain any constitutent $\mathfrak{D}_\tau$. Theorem 4 and corollary 2 now show that $\mathfrak{Z}$ has the kernel $\mathfrak{S}^*$. However, $\mathfrak{Z}$ was a (1–1)-representation. We thus find $\mathfrak{S}^* = \{1\}$, and corollary 3 shows that $\mathfrak{G}'$ is simple (or $\mathfrak{G}' = \{1\}$, if $g = p$).

Any doubly transitive permutation group $\mathfrak{G}$ of degree $p + 1$ possesses an irreducible (1–1)-representation of degree p. As condition (*) holds for any such $\mathfrak{G}$, we easily see that $\mathfrak{G}$ has the center $\{1\}$. It is, therefore, sufficient to treat irreducible linear groups with the center $\{1\}$ in a p-dimensional space. It follows here that $\mathfrak{S}^*$, considered as a linear group, must be reducible since the dimension does not divide the order s^*. Since $\mathfrak{S}^*$ is a normal subgroup, it splits into constituents of the same degree z. Then $z = 1$, and $\mathfrak{S}^*$ is abelian.

The last statement of theorem 6 follows from corollary 3.

Corollary 7. *If $\mathfrak{G}$ is a primitive irreducible linear group in a p-dimensional space and if $\mathfrak{G}$ has the center $\{1\}$, then $\mathfrak{G}'$ is simple.*

Proof: When $\mathfrak{G}$ is primitive, the normal abelian subgroup $\mathfrak{S}^*$ must lie in the center. We then have $\mathfrak{S}^* = \{1\}$ under our assumptions. Now corollary 3 gives the statement.

There is a well known theorem of Burnside which states that a transitive permutation group $\mathfrak{G}$ of degree p is either doubly transitive or it is metacyclic of the type (5). With the methods used here, this could be proved in the fol-

lowing manner. If $\mathfrak{G}$ is not doubly transitive, the permutation representation $\mathfrak{Z}$ splits into the 1-representation $\mathfrak{A}_1$ and at least two more constitutents not all of which can be of type $\mathfrak{A}_\rho$. Then one constituent at least is a $\mathfrak{C}^{(\nu)}$. Since the character is rational, all the conjugate $\mathfrak{C}^{(\nu)}$ appear, and we find

$$\mathfrak{Z} = \mathfrak{A}_1 + \sum_{\nu=0}^{t-1} \mathfrak{C}^{(\nu)}, \qquad c = (p-1)/t, \qquad t \geqq 2.$$

If $t > 2$, the group $\mathfrak{G}$ must have a normal subgroup of order p,[10] and therefore $\mathfrak{G}$ is of type (5). If $t = 2$, we have also to consider the case that $\mathfrak{G} \cong LF(2, p)$. It is not very difficult to exclude this last possibility. However, the proof may be omitted.

3. Conditions for the degrees a_ρ, b_σ, c

We now make use of the fact that the degrees a_ρ, b_σ, c (cf. (13)), divide the order $g = (p-1)p(1+np)/t$ of $\mathfrak{G}$. For the a_ρ, we certainly must have $(u_\rho p + 1) \mid (p-1)(1+np)$. If the sign δ has the value $+1$, the condition for c gives $(wp+1) \mid (p-1)(1+np)$. The question we have to treat then is this: When is

$$(up+1) \mid (p-1)(1+np)$$

(where $u = u_\rho$ or $u = w$)? Set $up + 1 = m_1 m_2$ with $m_1 \mid (p-1)$ and $m_2 \mid (1+np)$. Then $p \equiv 1$, $up + 1 \equiv 0 \pmod{m_1}$ and hence $u + 1 \equiv 0 \pmod{m_1}$. Further, $up + 1 \equiv 0$, $1 + np \equiv 0 \pmod{m_2}$, and this gives $(n-u)p \equiv 0$ and hence $n - u \equiv 0 \pmod{m_2}$. It now follows that $up + 1 = m_1 m_2$ divides $(u+1)(n-u)$. We may set

$$(u+1)(n-u) = h(up+1) \tag{18}$$

with an integral h. Then $h(up+1) \equiv 0 \pmod{u+1}$. Since $u \equiv -1 \pmod{u+1}$, this gives

$$(u+1) \mid h(p-1). \tag{19}$$

Assume first that $h = -h' < 0$. Then $(u+1)(u-n) = h'(up+1)$ which gives

$$u^2 = u(h'p + n - 1) + h' + n > u(h'p + n - 1).$$

Hence $u > h'p + n - 1$ while (19) yields $h'(p-1) \geqq u + 1$, i.e. $u \leqq h'p - h' - 1 \leqq h'p + n - 1$. This is a contradiction, the case $h < 0$ is impossible. From (18) it follows that $n = (hup + u^2 + u + h)/(u+1)$. We therefore have

Lemma 1. *Denote by $F(p, u, h)$ the rational function*

$$F(p, u, h) = \frac{hup + u^2 + u + h}{u+1} \tag{20}$$

[10] cf. [2], theorem 3.

of p, u and h. If $gt = (p-1)p(1+np)$ is divisible by a number $1+up$ where u is a non-negative integer, then there exists an integer $h \geqq 0$ such that

$$n = F(p, u, h) \tag{21}$$

and that $(u+1) \mid h(p-1)$.

In a similar fashion, we have to find the condition that $vp - 1$ divides $gt = (p-1)p(1+np)$ where $v > 0$, ($v = v_\sigma$, or $v = w$ for $\delta = -1$). We set $vp - 1 = m_1 m_2$ with $m_1 \mid (p-1)$ and $m_2 \mid (1+np)$ and derive $v - 1 \equiv 0 \pmod{m_1}$, $v + n \equiv 0 \pmod{m_2}$. Hence

$$(v-1)(v+n) = h(vp-1) \tag{22}$$

for some integer $h \geqq 0$. For fixed n, p, h, this is a quadratic equation for v. The second root $v' = (h-n)/v$ must be integral too. For $h = 0$, we have $v' = -n/v \leqq 0$. If $h \neq 0$, replace v by 1 in (22). The left side of (22) then vanishes while the right side is positive. This again gives $v' \leqq 0$.

Set $u = -v' = (n-h)/v$; then $u \geqq 0$. Since v' satisfies the equation (22), this equation remains correct when v is replaced by $-u$. We thus come back to (18). Since (18) implied (19) and (21), we have proved

LEMMA 2. *If $gt = (p-1)p(1+np)$ is divisible by $vp - 1$, where v is a positive integer, then there exists a non-negative integer h such that $u = (n-h)/v$ is integral and not negative, and that the relations hold*

$$n = F(p, u, h); \qquad (u+1) \mid h(p-1). \tag{23}$$

If $u = 0$, then $n = h$, and (22) gives $v = pn - n + 1$. However, it follows from (15) that $v_\sigma = pn - n + 1$ or $w = pn - n + 1$ are possible only when $v_\sigma = 1$ or $w = 1$, i.e. when $n = 0$. If $h = 0$, then $u = n$ and $v = 1$.

THEOREM 7. *If $\mathfrak{G}$ is a group satisfying the condition (*), then find all representations of n in the form $n = F(p, u^{(\nu)}, h^{(\nu)}) = (h^{(\nu)}u^{(\nu)}p + u^{(\nu)2} + u^{(\nu)} + h^{(\nu)})/(u^{(\nu)} + 1)$ with positive integers $u^{(\nu)}$, $h^{(\nu)}$. The degrees of the irreducible representations of $\mathfrak{G}$, as far as they are prime to p can only have some of the values*

$$a_\rho = 1, \qquad a_\rho = u^{(\nu)}p + 1, \qquad a_\rho = np + 1. \tag{24}$$

$$b_\sigma = p - 1, \qquad b_\sigma = v^{(\nu)}p - 1. \tag{25}$$

$$c = (np+1)/t, \qquad c = (u^{(\nu)}p + 1)/t, \qquad c = (p-1)/t, \qquad c = (v^{(\nu)}p - 1)/t \tag{26}$$

where $v^{(\nu)}$ is set equal to $(n - h^{(\nu)})/u^{(\nu)}$.

For $h > 0$ and variable u, we have $(\partial/\partial u)F(p, u, h) > 0$. Since we are only interested in solutions u, h of $n = F(p, u, h)$ with $1 \leqq u \leqq h(p-1) - 1$, we must have

$$F(p, 1, h) = \frac{hp + h + 2}{2} \leqq n \leqq F(p, h(p-1) - 1, h) = 2ph - h - 2. \tag{27}$$

This gives

THEOREM 8. *In theorem 7, only values $h = h^{(\nu)}$ have to be considered for which*

$$\frac{n+2}{2p-1} \leqq h \leqq \frac{2n-2}{p+1}. \tag{28}$$

To each $h^{(\nu)}$ there belongs at most one $u^{(\nu)}$, and this $u^{(\nu)}$ satisfies the conditions $u^{(\nu)} \mid (n - h^{(\nu)})$, $(u^{(\nu)} + 1) \mid h^{(\nu)}(p - 1)$.

The last remark follows from the fact that the equation $n = F(p, u, h)$ is equivalent to (18), and since $h < n$, we have one positive root u. Unless n has one of the following values

$$(29) \quad n = \begin{array}{llll} \frac{p+3}{2}, & \frac{2p+7}{3}, & \frac{3p+13}{4}, & \cdots \quad (h = 1; u = 1, 2, 3, \cdots) \\ p+2, & \frac{4p+8}{3}, & \frac{3p+7}{2}, & \cdots \quad (h = 2; u = 1, 2, 3, \cdots) \\ \frac{3p+5}{2}, & 2p+3, & \frac{9p+15}{4}, & \cdots \quad (h = 3; u = 1, 2, 3, \cdots) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots \end{array}$$

only the values

$$(30) \quad a_\rho = 1, \quad a_\rho = np + 1, \quad b_\sigma = p - 1, \quad c = (np + 1)/t, \quad c = (p - 1)/t$$

are possible in theorem 7.

As an example for theorems 7 and 8, we choose the case $t = 2$, $p = 11$, $n = 157$ for which $g = 95040$.[11] We find here $9 \leqq h \leqq 26$. Actually, only the values $h = 10, 12, 14, 17, 26$ give integral values for u. From (15), we obtain $u_\rho \leqq 28$, $v_\sigma \leqq 28$, $x_\tau \leqq 28$, $w \leqq 39$. We thus obtain as the possible values of the degrees a_ρ and b_σ :

$$a_\rho = 1, 12, 45, \text{or } 144; \qquad b_\sigma = 10, 32, 54, \text{or } 120,$$

while c has one of the values 5, 6, 16, 27, 60, 72, 160, 190. The sign $\delta = \pm 1$ is determined by $2c \equiv -\delta \pmod{11}$. The total number of degrees a_ρ and b_σ is five, and their values together with c must satisfy the equation (11). If we further assume that $\mathfrak{G}$ is simple, only one of the a_ρ has the value 1. On the basis of these remarks and using further properties of group representations, it is possible to compute the full table of group characters of $\mathfrak{G}$.

---

[11] This case is relatively complicated as n will appear five times in the table (29) for $p = 11$. The case has been chosen because there exists a simple group $\mathfrak{G}$ of order 95040, the five times transitive permutation group of degree 12 of Mathieu. It is possible to show that, for any simple group of order 95040 and $p = 11$, the condition (*) must hold and that t must have the value 2. The characters of the Mathieu groups have been obtained by Frobenius (Sitzungsberichte der Preussischen Akademie der Wissenschaften, Berlin 1903). Our result shows that any simple group of order 95040 has the same table of characters as the Mathieu group. This seems to make it appear highly probable that only one simple group of order 95040 exists.

4. Relations between the characters of $\mathfrak{G}$ and those of $\mathfrak{N} = \mathfrak{N}(\mathfrak{P})$

Every representation of $\mathfrak{G}$ defines a new representation of any given subgroup. We apply this for the subgroup $\mathfrak{N} = \mathfrak{N}(\mathfrak{P})$ defined by (5). In this manner, we can obtain some new information about the characters of $\mathfrak{G}$ from the formulas of section 1, and this will be needed later.

It is easy to find all irreducible characters of the group $\mathfrak{N}$ of order $pq = p(p-1)/t$.[12] Let ω be a primitive qth root of unity. We then have q linear characters ω_μ, $(\mu = 0, 1, 2, \cdots, q-1)$ defined by

$$\omega_\mu(Q^j) = \omega^{\mu j}, \qquad \omega_\mu(P^j) = 1. \tag{31}$$

Besides, we have t conjugate characters $Y^{(\nu)}$ of degree q. We only notice that

$$Y^{(\nu)}(Q^j) = 0 \qquad \text{for } j \not\equiv 0 \pmod{q}. \tag{32}$$

Set

$$\Omega = \omega_0 + \omega_1 + \cdots + \omega_{q-1}.$$

Then, we have

$$\Omega(1) = \Omega(P^i) = q, \qquad \Omega(Q^j) = 0 \qquad \text{for } j \not\equiv 0 \pmod{q}. \tag{33}$$

Every element N of $\mathfrak{N}$ which is not a power of P is conjugate to some Q^j and hence $\Omega(N)$ vanishes.

The expressions $A_\rho(N)$, $B_\sigma(N)$, $C^{(\nu)}(N)$, $D_\tau(N)$, (N in $\mathfrak{N}$), form (reducible) characters of $\mathfrak{N}$. Hence they can be expressed by the $\omega_\mu(N)$ and the $Y^{(\nu)}(N)$. From (33), (8) and (13), we obtain (N in the sums ranging over the elements of $\mathfrak{N}$)

$$\begin{aligned}
\sum A_\rho(N)\Omega(N) &= \sum_{i=0}^{p-1} A_\rho(P^i)\Omega(P^i) = q(a_\rho + p - 1) = qp(u_\rho + 1);\\
\sum B_\sigma(N)\Omega(N) &= \sum B_\sigma(P^i)\Omega(P^i) = q(b_\sigma - p + 1) = qp(v_\sigma - 1);\\
\sum C^{(\nu)}(N)\Omega(N) &= \sum C^{(\nu)}(P^i)\Omega(P^i) = q\left(c - \delta \sum_{i=1}^{p-1} \eta_i\right) = q(c + \delta q)\\
&= q(wp + \delta + \delta tq)/t = qp(w + \delta)/t;\\
\sum D_\tau(N)\Omega(N) &= \sum D_\tau(P^i)\Omega(P^i) = qd_\tau \qquad = pqx_\tau .
\end{aligned}$$

From the orthogonality relations for the characters of $\mathfrak{N}$, we now obtain

LEMMA 3. *If we consider the characters of $\mathfrak{G}$ only for elements N of the subgroup $\mathfrak{N}$, then $A_\rho(N)$ contains $u_\rho + 1$ of the $\omega_\mu(N)$, $B_\sigma(N)$ contains $v_\sigma - 1$ of the $\omega_\mu(N)$, $C^{(\nu)}(N)$ contains $(w + \delta)/t$ of the ω_μ, and $D_\tau(N)$ contains x_τ of the $\omega_\mu(N)$. The same $\omega_\mu(N)$ appear in the different $C^{(\nu)}(N)$, $\nu = 0, 1, 2, \cdots, t-1$.*

The last remark follows easily from the fact that $C(G)$ can be carried into

[12] The equations (31) and (32) can easily be derived from our general results applied to the case $n = 0$.

each $C^{(\nu)}(G)$ by a change in the choice of the primitive pth root of unity ϵ, because this change does not affect ω_μ. We also show

LEMMA 4. *We have*

$$(34)\quad \begin{cases} \sum_\rho a_\rho A_\rho(N) + \sum_\sigma b_\sigma B_\sigma(N) + c \sum_\nu C^{(\nu)}(N) + \sum_\tau d_\tau D_\tau(N) \\ \qquad\qquad = (1 + pn) \sum_{\mu=0}^{q-1} \omega_\mu(N) + \cdots . \end{cases}$$

$$(35)\quad \sum_\rho A_\rho(N) + \delta C(N) = \sum_\sigma B_\sigma(N) + \sum_{\mu=0}^{q-1} \omega_\mu(N) + \cdots .$$

$$(36)\quad \begin{aligned} &\sum_\rho u_\rho A_\rho(N) + \sum_\sigma v_\sigma B_\sigma(N) + wC(N) + \sum_\tau x_\tau D_\tau(N) \\ &\qquad\qquad = n \sum_{\mu=0}^{q-1} \omega_\mu(N) + \cdots . \end{aligned}$$

where the dots stand for a linear homogeneous combination of the $Y^{(\nu)}(N)$.

PROOF: The left side in (34) is the character $R(N)$ of the regular representation of $\mathfrak{G}$, and we have $R(1) = g$, $R(N) = 0$ for $N \neq 1$. Then (34) follows from the orthogonality relations and (31).

The expression

$$S(N) = \sum_\rho A_\rho(N) + \delta C(N) - \sum_\sigma B_\sigma(N)$$

can be written as a linear combination of the characters of $\mathfrak{N}$ with integral rational coefficients. From (8), (9) and (10), we obtain

$$\begin{aligned} \sum_{i=1}^{p-1} S(P^i) &= \sum_\rho (p-1) + \delta\left(-\delta \sum_{i=1}^{p-1} \eta_i\right) - \sum_\sigma (-p+1) \\ &= \alpha(p-1) + q + \beta(p-1) = pq; \end{aligned}$$

$$S(N) = 0 \quad \text{for} \quad N \neq P, P^2, \cdots, P^{p-1}; \qquad N \text{ in } \mathfrak{N}.$$

Then (31) gives

$$\sum S(N)\bar{\omega}_\mu(N) = pq$$

which shows that $S(N)$ contains ω_μ with the coefficient 1. Finally, (36) is obtained by subtracting (35) from (34) and dividing by p, taking into account that $C^{(\nu)}(N)$ and $C(N)$ contain the same ω_μ.

As a first application of the considerations of this section we prove

THEOREM 9.[13] *Let $\mathfrak{G}$ be a group which satisfies condition (*). If $\mathfrak{G}$ possesses an irreducible representation $\mathfrak{Z}$ of degree $p - 1$, then either the number t is even or the index of the commutator-group $\mathfrak{G}'$ in $\mathfrak{G}$ is even.*

PROOF: It follows from lemma 3 that the character $\zeta(N)$ of $\mathfrak{Z}(N)$ (N in $\mathfrak{N}$),

[13] If $p = 2$, then $t = 1 = p - 1$, and the theorem follows from Theorem 2.

contains only characters $Y^{(\nu)}(N)$ and no $\omega_\mu(N)$. We then must have $(p-1)/q = t$ such constituents $Y^{(\nu)}(N)$. It follows from (32) that $\mathfrak{Z}(Q)$ has the characteristic roots $1, \omega, \cdots, \omega^{q-1}$, each taken t times. Hence the determinant of $\mathfrak{Z}(Q)$ has the value $(-1)^{(q+1)t} = (-1)^{p-1+t} = (-1)^t$. The determinant of $\mathfrak{Z}(G)$, G in $\mathfrak{G}$, forms a representation of degree 1 of $\mathfrak{G}$. If t is odd, an even power of this representation will give the 1-representation of G. This implies that $(\mathfrak{G}:\mathfrak{G}')$ is even.

5. Proof of the main theorem

In what is to follow we shall assume that $\mathfrak{G}$ satisfies the following further condition

(**) *The commutator-subgroup of* $\mathfrak{G}$ *is equal to* $\mathfrak{G}$.

In the notation of section 2, we must have case II.[14] This shows that groups $\mathfrak{G}$ are excluded which contain a normal subgroup $\mathfrak{H}$ such that $\mathfrak{G}/\mathfrak{H}$ is metacyclic of order $p(p-1)/t$. Conversely, when $\mathfrak{G}$ is any group which satisfies the condition (*) and falls under case II, its commutator-subgroup $\mathfrak{G}'$ satisfies conditions (*) and (**). The number n is the same for $\mathfrak{G}$ and $\mathfrak{G}'$ and $\mathfrak{G}$ is obtained from $\mathfrak{G}'$ by a cyclic extension. We now prove

THEOREM 10. *Let* $\mathfrak{G}$ *be a group which contains an element* P *of prime order* p *which commutes only with its own powers and assume that* $\mathfrak{G}$ *is equal to its commutator-subgroup* $\mathfrak{G}'$. *If* $g = (p-1)p(1+np)/t$ *is the order of* $\mathfrak{G}$ *where* $1 + np$ *is the number of conjugate subgroups of order* p *in* $\mathfrak{G}$, *then the number* n *must be of the form*

$$n = \frac{puh + u^2 + u + h}{u + 1} \tag{37}$$

where u *and* h *are positive integers, except in the following two cases*

(a) $n = 1, t = 2$. *Here,* $\mathfrak{G} \cong LF(2, p), (p > 3)$.

(b) $n = \dfrac{p-3}{2}, t = \dfrac{p-1}{2}, p = 2^\mu + 1 > 3$ *a Fermat prime. Here,* $\mathfrak{G} \cong LF(2, p-1)$.

COROLLARY. *If* $n < (p+3)/2$, *then* $\mathfrak{G}$ *must be either of the type* (a) *or of the type* (b).

PROOF: Suppose that n is not representable in the form (37). Then the degrees of the irreducible representations of $\mathfrak{G}$ are either divisible by p or have one of the values (30). Because of condition (**) the degree 1 appears only once, say $a_1 = 1$. If t was odd, theorem 9 shows that the degree $p - 1$ does not appear. It follows from (11) that this is impossible. Hence

$$t \equiv 0 \pmod 2. \tag{38}$$

The degree $(p-1)/t$ is impossible for $t > 2$;[15] for $t = 2$ it occurs only in the case $\mathfrak{G} \cong LF(2, \text{p})$, i.e. in the case (a). Hence we may exclude this possibility.

---

[14] This implies that $p \neq 2$.

[15] If n is not of the form (37), it is not of the form $n = ump + u + m$ with positive integers u and m, since otherwise we could set $h = (u+1)m$ and would obtain a representation (37). Then corollary 6 (section 2) shows that $\mathfrak{G}'$ is simple. Because of conditions (**), $\mathfrak{G}$ is simple. Now, [2], theorems 3 and 4, can be applied.

We now see that we must have

$$
(39)\qquad \begin{aligned} &a_1 = 1, \qquad a_2 = a_3 = \cdots = a_\alpha = 1 + np, \\ &b_1 = b_2 = \cdots = b_\beta = p - 1, \qquad c = (pn + 1)/t \end{aligned}
$$

and the sign δ has the value $+1$. The values of α and β can be obtained from (9), (13), (14), and (39) which give

$$
(40)\qquad \alpha + \beta = (p - 1)/t = q,
$$

$$
(41)\qquad (\alpha - 1)n + \frac{n + 1}{t} = \beta.
$$

In particular, $n + 1$ is divisible by t; we set

$$
(42)\qquad n + 1 = st,
$$

and then obtain

$$
(43)\qquad 1 + np = 1 + (qt + 1)(st - 1) = tr
$$

where

$$
(44)\qquad r = qst + s - q = ps - q.
$$

By (43), the order g of $\mathfrak{G}$ can be written in the form

$$
(45)\qquad g = (p - 1)pr = qtpr.
$$

The next step is to show that r and $p - 1 = qt$ are relatively prime. Using (44), (42), (41), (40), we obtain successively

$$
s \equiv 0, \qquad n \equiv -1, \qquad (\alpha - 1)(-1) + s \equiv \beta, \qquad \alpha + \beta \equiv 0 \qquad (\text{mod } (r, q))
$$

whence it follows that $1 \equiv 0$, i.e. that $(r, q) = 1$. In a similar manner, we have

$$
s \equiv q, \qquad n \equiv -1, \qquad (\alpha - 1)(-1) + s \equiv \beta, \qquad \alpha + \beta \equiv s \quad (\text{mod } (r, t))
$$

which gives $s + 1 \equiv s$ (mod (r, t)), i.e. $(r, t) = 1$. Hence we have

$$
(46)\qquad (r, p - 1) = 1.
$$

From (45) and (46), it follows that for any prime l dividing $p - 1$ the characters B_σ are of highest kind.[16] This implies

$$
(47)\qquad B_\sigma(L) = 0
$$

for elements L of $\mathfrak{G}$ whose order is divisible by l. For the primes m dividing r the characters of degrees $1 + pn = rt$ and $(1 + pn)/t = r$ are of the highest kind. Hence

$$
(48)\qquad A_\rho(M) = 0 \quad \text{for} \quad \rho \neq 1, \qquad C^{(\nu)}(M) = 0
$$

[16] Cf. R. Brauer and C. Nesbitt, Annals of Mathematics, vol. 42, p. 556 (1941), Chapter II.

for elements M of $\mathfrak{G}$ whose order is divisible by m. Because of the assumption (*), the order of the elements L and M is not divisible by p, and hence the equation (10) holds for these elements. If an element G of $\mathfrak{G}$ would be an element L and an element M at the same time, then every term in (10) except $A_1(G) = 1$ would vanish, and this is impossible. Hence the elements of $\mathfrak{G}$ are distributed into four disjoint sets: (I) The 1-element, (II) the elements of order p, (III) the elements L whose order is divisible by at least one prime factor of $p - 1$, (IV) the elements M whose order is divisible by at least one prime factor of r.

Consider now the following element of the group ring Γ belonging to $\mathfrak{G}$

$$T = \sum_{j=1}^{q-1} Q^j. \tag{49}$$

We wish to show that (ρ now will always denote one of the values $2, 3, \cdots, \alpha$)

$$\begin{aligned} &A_1(T) = q - 1, \qquad A_\rho(T) = -(n + 1), \qquad B_\sigma(T) = 0, \\ &\qquad\qquad C^{(\nu)}(T) = -(n + 1)/t, \qquad D_\tau(T) = (q - 1)x_\tau . \end{aligned} \tag{50}$$

The proof of (50) can be obtained from the results of section 4. By (39) and (13) we have

$$u_1 = 0, \qquad u_2 = \cdots = u_\alpha = n, \qquad v_1 = \cdots = v_\beta = 1, \qquad w = n, \qquad \delta = 1.$$

Lemma 3 shows that $A_1(N)$ contains exactly one ω_μ, $A_\rho(N)$ with $\rho > 1$ contains exactly $n + 1$ of the ω_μ, $B(N)$ does not contain any ω_μ, $C(N)$ contains exactly $(n + 1)/t$ of the ω_μ, $D_\tau(N)$ contains x_τ of the ω_μ. Here, N is an element of $\mathfrak{N} = \mathfrak{N}(\mathfrak{P})$, cf. (5). It now follows from (35) that each of the ω_μ appears in one of the characters $A_1(N), A_2(N), \cdots, A_\alpha(N), C(N)$, while (36) shows that the ω_μ appearing in $A_\rho(N)$, $\rho > 1$, do not occur in any other character, and that the ω_μ appearing in $C(N)$ occur only in the $C^{(\nu)}(N)$. Since, obviously, $A_1(N) = 1 = \omega_0(N)$, this implies that the $A_\rho(N)$ with $\rho > 1$ and $C(N)$ contain only ω_μ for which $\mu \geqq 1$ whereas the D_τ contain only $\omega_0(N)$.

The elements Q^j belong to $\mathfrak{N}$. On account of (31) and (32), we have

$$\omega_0(T) = q - 1, \qquad \omega_\mu(T) = -1 \quad \text{for} \quad \mu \neq 0, \qquad Y^{(\nu)}(T) = 0.$$

The first formula (50) is obvious, as $A_1(N) = 1 = \omega_0(N)$. Since $A_\rho(N)$ for $\rho > 1$ is a sum of $n + 1$ terms ω_μ with $\mu > 0$ and of terms $Y^{(\nu)}$, we find $A_\rho(T) = -(n + 1)$. The remaining formulas (50) follow in the same manner using the facts that, apart from terms $Y^{(\nu)}$, the character $C^{(\nu)}(N)$ is a sum of exactly $(n + 1)/t$ terms ω_μ with $\mu > 0$, the character $D_\tau(N)$ contains only $x_\tau\omega_0$, and the characters $B_\sigma(N)$ do not contain any term except terms $Y^{(\nu)}$.

Let ζ range over all the characters $A_1, A_\rho, B_\sigma, C^{(\nu)}, D_\tau$ of $\mathfrak{G}$. If L again is an element whose order contains at least one prime factor of $p - 1$, then (47) and (50) give

$$\begin{aligned}\sum \zeta(T)\zeta(L) = (q-1) - (n+1)\sum_{\rho=2}^{\alpha} A_\rho(L) \\ - (n+1)C(L) + (q-1)\sum_{\tau} x_\tau D_\tau(L)\end{aligned} \tag{51}$$

since the t characters $C^{(\nu)}$ have the same value for the element L.

In order to compute these expressions in a different manner, we use the orthogonality relations for group characters. If ζ again ranges over all characters of $\mathfrak{G}$, we have

$$\begin{aligned}1 + \sum_{\rho=2}^{\alpha} A_\rho(G)A_\rho(H) + \sum_{\sigma} B_\sigma(G)B_\sigma(H) \\ + \sum_{\nu} C^{(\nu)}(G)C^{(\nu)}(H) + \sum_{\tau} D_\tau(G)D_\tau(H) \\ = \sum \zeta(G)\zeta(H) = n(G)\delta(G, H)\end{aligned} \tag{52}$$

where $n(G)$ is the order of the normalizer of $\mathfrak{G}$, and where $\delta(G, H)$ has the value 1 or 0 according as the elements G and H^{-1} of $\mathfrak{G}$ are or are not conjugate.

In particular, set $G = 1$ and $H = L$. The value $\zeta(1)$ is equal to the degree of the character and can be found from (39) and (13). Again we use (47) and the fact that $C^{(\nu)}(L) = C(L)$. Thus

$$1 + (1+np)\sum_{\rho} A_\rho(L) + (1+np)C(L) + p\sum_{\tau} x_\tau D_\tau(L) = 0.$$

Finally, (10) for $G = L$ reads

$$1 + \sum_{\rho} A_\rho(L) + C(L) = 0 \qquad (\rho \neq 1), \tag{53}$$

on account of (47). The last two equations give

$$\sum x_\tau D_\tau(L) = n \tag{54}$$

and, on combining (51), (53), (54), we obtain

$$\sum_{\zeta} \zeta(T)\zeta(L) = q - 1 + n + 1 + (q-1)n = q(n+1).$$

Substitute for T the value (49), and use again (52). This gives

$$\sum_{\zeta} \zeta(T)\zeta(L) = \sum_{\zeta}\sum_{j=1}^{q-1} \zeta(Q^j)\zeta(L) = \sum_{j} n(Q^j)\delta(Q^j, L).$$

Hence

$$\sum_{j=1}^{q-1} n(Q^j)\delta(Q^j, L) = q(n+1). \tag{55}$$

The equation (55) shows that not all $\delta(Q^j, L)$ can vanish. Hence L is conjugate to some power of Q

$$L \sim Q^\nu \qquad (1 \leqq \nu \leqq q-1). \tag{56}$$

Since L is any element whose order is divisible by at least one prime factor of $p - 1 = qt$, and since Q has the order q, this proves that every prime factor of t divides q. Moreover, the formula (55) can be applied for $L = Q^i$, $i \not\equiv 0 \pmod q$. The left hand side is obviously the order of the normalizer of the cyclic group $\{Q^i\}$, and if this order is denoted by $N\{Q^i\}$, we have

$$N\{Q^i\} = q(n + 1). \tag{57}$$

As the order of a subgroup, $q(n + 1) = qst$ (cf. (42)) must divide the order $g = qtpr$ (cf. (45)) of $\mathfrak{G}$. This gives $s \mid pr$. If $p \mid s$, (42) and (41) would imply that $\beta \geqq p$, and this contradicts (40). Hence $s \mid r$. Now (44) shows that s divides q and is, therefore, a common divisor of r and $qt = p - 1$. By (46), we have $s = 1$.

Now (42), (44), and (45) become

$$n = t - 1, \qquad r = p - q, \qquad g = (p - 1)p(p - q) \tag{58}$$

while (40) and (41) yield

$$\alpha t - t + 2 = q = (p - 1)/t. \tag{59}$$

As we have seen, every prime divisor of t must divide q. Because of (59), t must be a power of 2, say

$$t = 2^{\mu-1}. \tag{60}$$

It follows from (38) that $\mu \geqq 2$.

Consider first the case $\mu = 2$, i.e. $t = 2$. Here, $n = 1$, $q = (p - 1)/2$ and $g = p(p - 1)(p + 1)/2$. From (59) we obtain $\alpha = (p - 1)/4$. But then (40) shows that $\beta = (p - 1)/4$. We have therefore, one character of degree 1, $(p - 5)/4$ characters of degree $p + 1$, $(p - 1)/4$ characters of degree $p - 1$, two characters of degree $(p + 1)/2$, and one character of degree p. The last fact follows from (15) which now reads

$$\sum x_\tau^2 = \frac{p}{2} - \frac{p-5}{4} - \frac{p-1}{4} - \frac{1}{2} = 1.$$

The method applied in an earlier paper[17] now gives $\mathfrak{G} \cong LF(2, p)$.

Assume now $\mu > 2$, i.e. $t \geqq 4$. As shown in (56), all elements L of an order 2^λ are conjugate to a power of Q. But (59) shows for $t \equiv 0 \pmod 4$ that $q \not\equiv 0 \pmod 4$. Hence no elements of order 4 exist, and a 2-Sylow-subgroup $\mathfrak{P}$ of $\mathfrak{G}$ can contain only elements of order 1 and 2. This implies that $\mathfrak{P}$ is an abelian group of type $(2, 2, \cdots, 2)$.

Let $\mathfrak{T}_i$ be the normalizer of the group $\{Q^i\}$, $i \not\equiv 0 \pmod q$. Every $\mathfrak{T}_i$ is contained in $\mathfrak{T} = \mathfrak{T}_1$, but since (57) shows that all $\mathfrak{T}_i$ have the same order $q(n + 1) = qt = p - 1$, we have

$$\mathfrak{T} = \mathfrak{T}_1 = \cdots = \mathfrak{T}_{q-1}.$$

[17] [2], sections 9 and 10.

Assume now that $q \neq 2$. Since we have $q \not\equiv 0 \pmod 4$, the number q will contain an odd prime divisor l. We set $L_0 = Q^{q/l}$. If X is an element of order 2 in $\mathfrak{T}$, we have

$$X^{-1}L_0X = L_0^j$$

for some j. But X^2 will commute with L_0 which gives $j^2 \equiv 1 \pmod l$, i.e. $j \equiv \pm 1 \pmod l$.

The order $p - 1 = qt$ of $\mathfrak{T}$ is divisible by 8. Because of the structure of the 2-Sylow subgroup $\mathfrak{L}$ of $\mathfrak{G}$, it follows that $\mathfrak{T}$ contains an abelian subgroup of type $(2, 2, 2)$. This implies that we have at least three elements X of order 2 in $\mathfrak{T}$ for which $j \equiv 1 \pmod l$.

If $j \equiv 1 \pmod l$, X and L_0 commute and $L = XL_0$ is an element of order $2l$. By (56), XL_0 then is conjugate in $\mathfrak{G}$ to a power of Q, say

$$T^{-1}XL_0T^{-1} = Q^{\rho}. \tag{61}$$

Hence $T^{-1}L_0^2T = Q^{2\rho}$, i.e. $T^{-1}Q^{2q/l}T = Q^{2\rho}$. This shows that T belongs to the normalizer of $Q^{2q/l} \neq 1$, and therefore T belongs to $\mathfrak{T}$. Moreover, (61) implies that $T^{-1}X^lT = Q^{\rho l}$. Hence $Q^{\rho l}$ has order 2, i.e. $Q^{\rho l}$ must be equal to the only power $Q^{q/2}$ of Q of order 2. Since T transforms $Q^{q/2}$ into itself, we have $X = X^l = T^{-1}Q^{q/2}T = Q^{q/2}$. This gives a contradiction because we had shown that we have at least three different elements X of order 2 in $\mathfrak{T}$ for which $j \equiv 1$. It follows that $q = 2$ and then (58) and (60) give

$$p = qt + 1 = 2^{\mu} + 1, \qquad g = p(p-1)(p-2), \qquad t = \frac{p-1}{2}, \qquad n = \frac{p-3}{2}. \tag{62}$$

In particular, p must be a Fermat prime. Furthermore, (59) and (40) give

$$\alpha = 1, \qquad \beta = 1.$$

Hence $\mathfrak{G}$ has one representation A_1 of degree 1, one representation of degree $p - 1$, and $(p - 1)/2$ representations of degree $(1 + np)/t = p - 2$. The degree of all other irreducible representations is divisible by p.

The 2-Sylow subgroup $\mathfrak{L}$ must have the order $p - 1 = 2^{\mu}$. We may assume that $\mathfrak{L}$ contains Q. From (56) it follows that each of the $2^{\mu} - 1$ elements $L \neq 1$ of $\mathfrak{L}$ is conjugate to Q in $\mathfrak{G}$. Then L must be conjugate to Q in the normalizer $\mathfrak{N}(\mathfrak{L})$ of the abelian group $\mathfrak{L}$. Conversely, every element of $\mathfrak{N}(\mathfrak{L})$ transforms Q into an element $L \neq 1$ of $\mathfrak{L}$. But (57) for $i = 1$ and (62) give

$$N(Q) = q(n + 1) = 2(p - 1)/2 = p - 1 = 2^{\mu}.$$

Hence only the elements of $\mathfrak{L}$ will commute with Q. This shows that $\mathfrak{L}$ has the index $2^{\mu} - 1$ in $\mathfrak{N}(\mathfrak{L})$. Consequently, $\mathfrak{N}(\mathfrak{L})$ has the order $2^{\mu}(2^{\mu} - 1) = (p - 1)(p - 2)$ i.e. $\mathfrak{N}(\mathfrak{L})$ has the index p in $\mathfrak{G}$.

It now follows that $\mathfrak{G}$ possesses a permutation representation of degree p. If Π is the corresponding character, then Π contains A_1 exactly once. Since $p > 3$, Π cannot contain a character $C^{(\nu)}$ of degree $p - 2$, and it cannot contain any character D_{τ}. Hence we have

$$\Pi(G) = A_1(G) + B_1(G) = 1 + B_1(G).$$

From (8), (10), (47), and (48), we obtain $B_1(P^i) = -1$ for $i \not\equiv 0 \pmod p$, $B_1(L) = 0$, $B_1(M) = A_1(M) + C(M) = 1$.

Hence

(63) $\Pi(1) = p, \quad \Pi(P^i) = 0$ for $i \not\equiv 0 \pmod p, \quad \Pi(L) = 1, \quad \Pi(M) = 2$

where L and M have the same significance as in (47), (48). However, $\Pi(G)$ equals the number of letters not altered by the permutation representing G. Then (63) shows that we have a (1–1)-representation since $\Pi(G) = p$ only for $G = 1$. The subgroup leaving three letters fixed has the order 1, i.e. its index in $\mathfrak{G}$ is $p(p-1)(p-2)$. This implies that $\mathfrak{G}$ is three times transitive. From a theorem of Zassenhaus,[18] it now follows that $\mathfrak{G} \cong LF(2, p-1)$ and this finishes the proof of theorem 10.

6. Simple groups of order $(p-1)\,p\,(1+mp)/\tau$

We now drop the assumption (*) and propose to prove the following theorem

THEOREM 11. *Let $\mathfrak{G}$ be any non-cyclic simple group of order*

$$g = (p-1)p(1+mp)/\tau$$

where p is a prime, and where τ and m are any non-negative integers such that τ divides $p - 1$. If m does not possess a representation of the form $m = (puh + u^2 + u + h)/(u+1)$ with positive integers u, h, in particular if $m < (p+3)/2$, then $\mathfrak{G}$ is either isomorphic to $LF(2, p)$ or to $LF(2, p-1)$, and in the second case p must be a Fermat prime, $p = 2^\mu + 1$, $(p > 3)$.

PROOF. Let $1 + np$ be the number of conjugate subgroups of order p in $\mathfrak{G}$. If $\mathfrak{P} = \{P\}$ is one of these Sylow-subgroups, then $g/(1+np)$ is the order of the normalizer of $\mathfrak{N}(\mathfrak{P})$. If the order of the normalizer (centralizer) of P is v, and if t classes of conjugate elements of $\mathfrak{G}$ contain elements of order p, we have

(64) $$g = (p-1)p(1+mp)/\tau = (p-1)pv(1+np)/t.$$

The number n here is positive since otherwise $\mathfrak{P}$ would be a normal subgroup of $\mathfrak{G}$. The number $1 + np$ of conjugate Sylow subgroups divides g and hence it divides $g\tau = (p-1)p(1+mp)$. Then lemma 1 shows that there exists an integer $h \geqq 0$ such that

(65) $$m = F(p, n, h).$$

Since it was assumed that m did not have a representation $m = F(p, u, h)$ with positive integers u and h, we must have $h = 0$ in (65). Then n becomes equal to m, and (64) gives

(66) $$n = m, \qquad t = v\tau.$$

Consider now the irreducible representations of $\mathfrak{G}$ which belong to the first

[18] H. Zassenhaus, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 11, p. 17 (1935).

p-block. For this block we have corresponding results[19] as for the A_ρ, B_σ, $C^{(\nu)}$ in section 1. There exists, however, other representations of a degree prime to p when $v > 1$, and (12) and (15) will no longer hold.

The degrees $a_\rho = pu_\rho + 1$ and $b_\sigma = pv_\sigma - 1$ in the first p-block must divide $g\tau = (p - 1)p(1 + pm)$. Then lemmas 1 and 2 show that we have $a_\rho = 1$ or $a_\rho = 1 + pm$ and $b_\sigma = p - 1$, cf. (30). Since $\mathfrak{G}$ is simple, only one a_ρ has the value 1, say $a_1 = 1$.

If $v = 1$, the group $\mathfrak{G}$ satisfies the assumption (*), and theorem 10 implies theorem 11 in this case. We therefore may assume that $v > 1$. Suppose now that $V \neq 1$ is a p-regular element of $\mathfrak{N}(P)$. If $\beta \neq 0$, it follows easily from (8, II) that $\mathfrak{B}_\sigma(P)$ has the characteristic roots $\epsilon, \epsilon^2, \cdots, \epsilon^{p-1}$ where ϵ is a primitive pth root of unity. We may arrange the characteristic roots $x_1, x_2, \cdots, x_{p-1}$ of $\mathfrak{B}_\sigma(V)$ in such a manner that P^iV has the characteristic roots $x_\mu \epsilon^{i\mu}$, ($\mu = 1, 2, \cdots, p - 1$). However, $\mathfrak{B}_\sigma(P^iV) = -1$ for $i \not\equiv 0 \pmod p$.[20] This gives

$$\sum_{\mu=1}^{p-1} \epsilon^{\mu i} x_\mu = -1 = \sum_{\mu=1}^{p-1} \epsilon^{\mu i} \qquad (i = 1, 2, \cdots, p - 1). \tag{67}$$

The x_μ are themselves roots of unity, of an exponent prime to p. Since $\epsilon, \epsilon^2, \cdots, \epsilon^{p-1}$ must be linearly independent over the field generated by the x_μ, (67) implies $x_\mu = 1$. But then $\mathfrak{B}_\sigma(V) = I$, and since $\mathfrak{G}$ was simple, we have $V = 1$ which gives a contradiction. Hence we have $\beta = 0$.

It now follows from (9) that

$$a_1 = 1, \qquad a_2 = \cdots = a_q = pm + 1;$$

and (11) gives $\delta = -1$ and

$$c = 1 + (q - 1)(pm + 1). \tag{68}$$

This shows that $1 + pm$ and c are relatively prime. The degree c divides g, and (64) now implies that c divides $(p - 1)/\tau$. However, since $c \neq 1$ and $n = m \neq 0$, the equation (68) yields $c > p > (p - 1)/\tau$. We have a contradiction, and theorem 11 is proved.

7. Examples of groups which satisfy the condition (*)

First, let $\mathfrak{G}$ be a transitive permutation group on p letters, where p is a prime. The order g of $\mathfrak{G}$ then is divisible by p, and $\mathfrak{G}$ contains elements P of order p. Each such element P is represented by a simple cycle of length p. It now follows easily that P commutes only with its own powers; i.e. $\mathfrak{G}$ satisfies the condition (*). In a similar manner, we can show that a doubly transitive group of degree $p - 1$ satisfies the condition (*).

We next consider irreducible groups $\mathfrak{G}$ of linear transformations of a p-dimensional vector space[21] where p again is a prime. Assume that $\mathfrak{G}$ is of finite

[19] cf. [1], theorems 4 and 11.
[20] cf. [1], theorems 4 and 11.
[21] cf. footnote 2.

order g and that its center consists of the unit element only. The order g is divisible by the degree p of the irreducible group. Let $\mathfrak{P}$ be a Sylow-subgroup of $\mathfrak{G}$, and let P_0 be an invariant element of $\mathfrak{P}$ which is different from 1. Then P_0 cannot be a scalar multiple of the unit matrix, since it does not belong to the center $\{1\}$ of $\mathfrak{G}$. But P_0 commutes with every element of $\mathfrak{P}$, and Schur's lemma implies that the linear group $\mathfrak{P}$ is reducible. The degree of every irreducible constituent of $\mathfrak{P}$ must be 1, since it divides the order of $\mathfrak{P}$. Hence $\mathfrak{P}$ can be taken as a set of diagonal matrices, i.e. $\mathfrak{P}$ is an abelian group. We now prove

LEMMA 5.[22] *Let $\mathfrak{G}$ be a group of order $g = p^a g^*$ with $(p, g^*) = 1$, and assume that $\mathfrak{G}$ does not contain invariant elements of order p, and that the Sylow-subgroup $\mathfrak{P}$ of order p^a in $\mathfrak{G}$ is abelian. If $\mathfrak{Z}$ is an irreducible* (1–1) *representation of $\mathfrak{G}$ of degree p^μ, then $\mu = a$.*

PROOF. Let ζ be the character of $\mathfrak{Z}$. If G is an element of $\mathfrak{G}$ which has exactly j conjugate elements then it is well known that $j\zeta(G)/\zeta(1) = j\zeta(G)/p^\mu$ is an algebraic integer. Since $\mathfrak{P}$ is abelian, the number j is prime to p when G lies in $\mathfrak{P}$. Hence $\zeta(G) \equiv 0 \pmod{p^\mu}$ for G in $\mathfrak{P}$. On the other hand, $\zeta(G)$ is a sum of p^μ roots of unity. If $G \neq 1$, not all these roots of unity can be equal. Hence the integer $\zeta(G)/p^\mu$ and all its algebraic conjugates are smaller than 1 in absolute value which implies $\zeta(G) = 0$ for every $G \neq 1$ in $\mathfrak{P}$.[23] Then the character ζ is of the highest kind,[24] i.e. $p^\mu = \zeta(1) \equiv 0 \pmod{p^a}$ which yield $\mu = a$, as was stated.

LEMMA 6. *Let $\mathfrak{G}$ be a group of order $g = p^a g^*$ with $(p, g^*) = 1$. If the center of $\mathfrak{G}$ consists of the unit element only, and if $\mathfrak{G}$ has a* (1–1)*-representation $\mathfrak{Z}$ of degree p^a, then the centralizer of a Sylow subgroup $\mathfrak{P}$ of order p^a is contained in $\mathfrak{P}$.*

PROOF. The character ζ of $\mathfrak{Z}$ has the values[25]

$$\zeta(P) = \begin{cases} p^a & P = 1 \\ 0 & P \text{ in } \mathfrak{P}, \quad P \neq 1. \end{cases}$$

Hence $\mathfrak{Z}(P)$ is the regular representation of $\mathfrak{P}$; we may assume that it breaks up into the distinct irreducible representations $\mathfrak{F}_\mu$ of $\mathfrak{P}$, each $\mathfrak{F}_\mu$ appearing f_μ times where f_μ is the degree of $\mathfrak{F}_\mu$. We then have

$$\mathfrak{Z}(P) = \begin{pmatrix} \ddots & & 0 \\ & f_\mu \times F_\mu & \\ 0 & & \ddots \end{pmatrix}, \qquad F_\mu = \mathfrak{F}_\mu(P).$$

---

[22] The following lemmas 5 and 6 are proved here in a more general form than necessary for our purpose. However, in the form given here, they can also be used in other connections.

[23] For this argument, cf. W. Burnside, Proceedings of the London Mathematical Society (2) vol. 1, p. 388–392 (1904).

[24] cf. theorem 10 of the paper mentioned in footnote 16.

[25] cf. footnote 16.

Let V be an element of the centralizer $\mathfrak{C}$ of $\mathfrak{P}$ such that the order v of V is prime to p. Then $\mathfrak{Z}(V)$ will commute with $\mathfrak{Z}(P)$. It follows that $\mathfrak{Z}(V)$ is of the form

$$\mathfrak{Z}(V) = \begin{pmatrix} \ddots & & 0 \\ & T_\mu \times I_\mu & \\ 0 & & \ddots \end{pmatrix}$$

where T_μ is a matrix of degree f_μ and I_μ is the unit matrix of degree f_μ. Then $\mathfrak{Z}(V^i P)$ breaks up completely into the matrices $T_\mu^i \times F_\mu$. If $\tau_\mu^{(i)}$ is the trace of T_μ^i and if $\theta_\mu(P)$ is the character of $F_\mu(P)$, we find

$$\zeta(V^i P) = \sum \tau_\mu^{(i)} \theta_\mu(P).$$

Since ζ is of the highest kind, we have

$$\sum_\mu \tau_\mu^{(i)} \Theta_\mu(P) = 0 \qquad \text{for } P \text{ in } \mathfrak{P}, \quad P \neq 1. \tag{70}$$

Set $\sum \tau_\mu^{(i)} f_\mu = \tau^{(i)}$ where the sum extends over all values of μ. Using the orthogonality relations for the characters of $\mathfrak{P}$, we derive from (70) the equation

$$p^a \tau_\nu^{(i)} = \sum_P \sum_\mu \tau_\mu^{(i)} \Theta_\mu(P) \Theta_\nu(P^{-1}) = \sum_\mu \tau_\mu^{(i)} f_\mu f_\nu = f_\nu \tau^{(i)}.$$

This shows that the matrices T_ν^i / f_ν have the same trace for all values of ν. If $\mathfrak{F}_1$ is the 1-representation of $\mathfrak{P}$, then T_1 is a vth root of unity λ where v is the order of V. Hence

$$\operatorname{tr}(T_\nu^i) = \tau_\nu^{(i)} = f_\nu \tau^{(i)}/p^a = f_\nu \tau_1^{(i)} = f_\nu \lambda^i.$$

The mapping $V^i \to T_\nu^i$ defines a representation of the group $\{V\}$. Since its character is identical with the character of the representation $V^i \to \lambda^i I_\nu$, it follows easily that $T_\nu^i = \lambda^i I_\nu$. Hence $\mathfrak{Z}(V) = \lambda I$. This is impossible for $V \neq 1$, because $\mathfrak{Z}$ was a (1–1)-representation, and $\mathfrak{G}$ did not contain any invariant elements except 1.

Hence the centralizer $\mathfrak{C}$ of $\mathfrak{P}$ cannot contain elements of an order prime to p. Consequently, the order of $\mathfrak{C}$ itself is a power of p. Since $\mathfrak{C}$ and $\mathfrak{P}$ generate a p-group contained in $\mathfrak{G}$, we have $\mathfrak{C} \subseteq \mathfrak{P}$ and this proves lemma 6.

Returning to irreducible linear groups $\mathfrak{G}$ of degree p whose center consists only of the unit element, it follows from lemma 5 that the order g contains p to the first power only. If P is an element of order p, then lemma 6 shows that $\{P\}$ is the centralizer of P. Hence we see

THEOREM 12. *If p is a prime, then the transitive permutation groups p, the doubly transitive permutation groups of degree $p + 1$, and the irreducible linear groups in p dimensions with the center $\{1\}$ satisfy the condition* (*).

THE INSTITUTE FOR ADVANCED STUDY.

ANNALS OF MATHEMATICS
Vol. 44, No. 1, January, 1943

ON THE CONVERGENCE OF CONTINUED FRACTIONS TO MEROMORPHIC FUNCTIONS*

BY WALTER LEIGHTON AND W. J. THRON

(Received August 28, 1942)

1. Introduction

In this paper we give a new set of convergence criteria for continued fractions of the form

$$\frac{1}{1} + \frac{a_2}{1} + \frac{a_3}{1} + \cdots, \tag{1}$$

where the a_n are complex numbers. The fundamental results are contained in Theorem 3. Application is made to continued fractions of the form

$$1 + \frac{a_1 x}{1} + \frac{a_2 x}{1} + \cdots, \tag{2}$$

where x is a complex variable. The basic algorithm in §3 is equivalent to one introduced by Euler [1, p. 206]. A direct proof is included since it is very brief.

2. General concepts

The continued fraction

$$b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \cdots,$$

where the a_n and b_n are complex numbers, is said to converge or diverge according as the series

$$\frac{A_0}{B_0} + \left(\frac{A_1}{B_1} - \frac{A_0}{B_0}\right) + \left(\frac{A_2}{B_2} - \frac{A_1}{B_1}\right) + \cdots \tag{3}$$

converges or diverges. If the series converges to a value v, the continued fraction is said to converge to v. The numbers A_n and B_n are defined by the recursion relations

$$\begin{aligned} A_0 &= b_0, \ A_1 = b_0 b_1 + a_1, \\ B_0 &= 1, \ B_1 = b_1, \\ A_n &= b_n A_{n-1} + a_n A_{n-2}, \\ B_n &= b_n B_{n-1} + a_n B_{n-2}. \end{aligned} \tag{4}$$

The ratio A_n/B_n is called the n^th approximant of the continued fraction. Finally, if it is known that the series either converges or, at worst, diverges to ∞, the

* Presented to The American Mathematical Society, Sept. 8, 1942.

continued fraction is said to *converge at least in the wider sense.* If the numbers a_n and b_n are functions of a variable x, and if for x in a prescribed set the series (3) converges uniformly with respect to x, the continued fraction is said to converge uniformly with respect to x in this set.

3. An algorithm

It is well known that the quantity

$$D_n = \frac{A_n}{B_n} - \frac{A_{n-1}}{B_{n-1}} \qquad (n = 1, 2, 3, \cdots),$$

is given by the relation

$$D_n = \frac{(-1)^{n-1} a_1 a_2 \cdots a_n}{B_{n-1} B_n}, \tag{5}$$

at least when $B_{n-1}B_n \neq 0$. The numbers D_n assume a particularly simple form for the continued fraction

$$\frac{1}{1} + \frac{z_2 - 1}{1} + \frac{z_2(z_3 - 1)}{1} + \frac{z_3(z_4 - 1)}{1} + \cdots. \tag{6}$$

LEMMA. *For the continued fraction* (6) *the quantity D_n is given by the formula*

$$D_n = (-1)^{n-1} \frac{(z_2 - 1)(z_3 - 1) \cdots (z_n - 1)}{z_2 z_3 \cdots z_n} \qquad (n \geqq 2).$$

This result will follow from (5) once it has been shown that

$$B_n = z_2 z_3 \cdots z_n \qquad (n \geqq 2). \tag{5$'$}$$

The proof of (5)′ is by induction. This formula may be verified readily for $n = 2$ and for $n = 3$. Suppose it holds for $n = k - 1$ and for $n = k$. By the last relation (4)

$$B_{k+1} = B_k + z_k(z_{k+1} - 1)B_{k-1}.$$

The proof can now be completed by applying the induction hypothesis to B_k and to B_{k-1}.

It is an immediate consequence of the lemma that for the continued fraction (6) the series (3) takes the form

$$\sum_{n=1}^{\infty} D_n = 1 + \sum_{n=2}^{\infty} (-1)^{n-1} \left(1 - \frac{1}{z_2}\right)\left(1 - \frac{1}{z_3}\right) \cdots \left(1 - \frac{1}{z_n}\right).$$

THEOREM 1. *The continued fraction* (6) *converges if the numbers z_n $(n \geqq 2)$ satisfy the relation*

$$\left| z_{2n} + \frac{1}{h^2 - 1} \right| > \frac{h}{h^2 - 1} + e \tag{7}$$

and

$$\left| z_{2n+1} - \frac{h^2}{h^2-1} \right| < \frac{h}{h^2-1} - d, \tag{8}$$

where h is any number > 1, and e and d are arbitrary positive numbers.

If relations (7) and (8) hold, positive numbers e' and d' can be found such that

$$\left| 1 - \frac{1}{z_{2n}} \right| < h - e'$$

and

$$\left| 1 - \frac{1}{z_{2n+1}} \right| < \frac{1}{h + d'}.$$

Under these conditions

$$| D_n | < \frac{(h - e')^{[n/2]}}{(h + d')^{[(n-1)/2]}}.$$

The series (3) is therefore majorized by the convergent series

$$1 + (h - e') + \frac{(h - e')}{(h + d')} + \frac{(h - e')^2}{(h + d')} + \cdots ; \tag{9}$$

hence, the series (3) and the continued fraction (6) converge. The proof is complete.

THEOREM 2. *If the quantities z_n are functions of a complex variable x and if for all values of x contained in a certain region D, the functions $z_{2n}(x)$ and $z_{2n+1}(x)$ satisfy relations* (7) *and* (8) *respectively, where h, e and d are independent of x, the continued fraction* (6) *converges uniformly for all x in D. Further, if the functions $z_n(x)$ are analytic for all x in D, the continued fraction* (7) *converges to a function that is analytic for all x in D.*

Under the conditions of the theorem, the series (9) is independent of x and majorizes the series (3) for all x in D. Thus the series (3) converges uniformly. The partial sums of the series (3) are rational functions of the quantities z_n and hence analytic functions of x. Weierstrass' well-known theorem then insures the validity of the second part of the theorem.

4. Convergence criteria

The results of the preceding section will now be applied to continued fractions (1). It is our aim to find conditions on the numbers a_n such that for any given sequence $\{a_n\}$ satisfying these conditions a sequence $\{z_n\}$ is uniquely determined by the system of equations

$$\begin{aligned} a_2 &= z_2 - 1, \\ a_{2n} &= z_{2n-1}(z_{2n} - 1), \\ a_{2n+1} &= z_{2n}(z_{2n+1} - 1), \end{aligned} \tag{10}$$

in such a way that the numbers z_n satisfy the conditions of Theorem 1. Conditions of this kind on the numbers a_n will clearly entail convergence of the continued fraction (1).

We proceed with the following definitions:

$$\text{(11)}\quad \begin{aligned} &\text{(i)} \quad z \in Z_1, \quad \text{if} \quad \left| z - \frac{h^2}{h^2 - 1} \right| < \frac{h}{h^2 - 1}, \\ &\text{(ii)} \quad z \in Z_2, \quad \text{if} \quad \left| z + \frac{1}{h^2 - 1} \right| > \frac{h}{h^2 - 1}, \\ &\text{(iii)} \quad w \in W_1, \quad \text{if} \quad \left| w - \frac{1}{h^2 - 1} \right| < \frac{h}{h^2 - 1}, \\ &\text{(iv)} \quad w \in W_2, \quad \text{if} \quad \left| w + \frac{h^2}{h^2 - 1} \right| > \frac{h}{h^2 - 1}. \end{aligned}$$

It follows that $z \in Z_1$ implies $z - 1 \in W_1$ and that $z \in Z_2$ implies $z - 1 \in W_2$, and conversely.

We shall determine regions $A_1(h)$ and $A_2(h)$ with the property that if $a_{2n+1} \in A_1(h)$ and $a_{2n} \in A_2(h)$, the numbers $z_n(\neq 0)$ defined by equations (10) will belong to Z_1, when n is odd, and to Z_2, when n is even.

To this end let us denote by $z \cdot W$ the set of points containing all products zw, where z is fixed and w ranges over the set W. By $D[z \cdot W_2]$ we shall denote the point set intersection of all such sets $z \cdot W_2$ as z is allowed to vary over the region Z_1. It follows that a is an element of $D[z \cdot W_2]$ if and only if a/z is an element of W_2 for all $z \in Z_1$.

The region $A_2(h)$ may be taken as $D[z \cdot W_2]$, since for any $a \in A_2(h)$ and $z \in Z_1$ there exists a $z' \in Z_2$ (i.e., $z' - 1 \in W_2$) such that $a = z(z' - 1)$.

Similarly the region $A_1(h)$ may be taken as $D[z \cdot W_1]$ where z assumes all values of the set X_2. As $z = 1 \in Z_1$ it is clear that the set $A_2(h)$ is contained in the set W_2.

If now a sequence $\{a_n\}$ with $a_{2n} \in A_2(h)$ and $a_{2n+1} \in A_1(h)$ is given, a sequence $\{z_n\}$ can be uniquely determined ($z_n \neq 0$) satisfying the system of equations (10). If further the numbers a_{2n} and a_{2n+1} are bounded away from the boundaries of their respective regions, then an e and a d can be found such that the corresponding z_n satisfy conditions (7) and (8) respectively.

From the definition of the set $A_2(h)$ it follows that it is that part of the complex plane which remains when all circular regions $z \cdot C$ are deleted, where z is allowed to vary over the closure of the region Z_1 and C is the closed circular region

$$\left| z + \frac{h^2}{h^2 - 1} \right| \leqq \frac{h}{h^2 - 1}.$$

The boundary of the region $A_2(h)$ can be determined as follows. Consider the products zc where z and c are arbitrary values from the regions Z_1 and C

respectively. These products lie outside $A_2(h)$ and can lie on the boundary of $A_2(h)$ only if both z and c lie on the boundaries of their respective regions. Let us fix the sum of the arguments of z and c. The products then lie on a ray passing through the origin, and only those products will lie on the boundary of $A_2(h)$ for which the absolute value of the product zc, for a fixed arg zc, is maximized and minimized respectively.

We now note that the circles that form the boundaries of Z_1 and C have the same radius and have centers on the real axis with the x coordinates $h^2/(h^2 - 1)$ and $-h^2/(h^2 - 1)$ respectively.

We further note that for any fixed arg z there are two points z on the circle bounding Z_1. A similar remark is true for C. Let $|z_1| < |z_2|$, $\arg z_1 = \arg z_2$ and $|c_1| < |c_2|$, $\arg c_1 = \arg c_2$. It is clear that the minimum and maximum

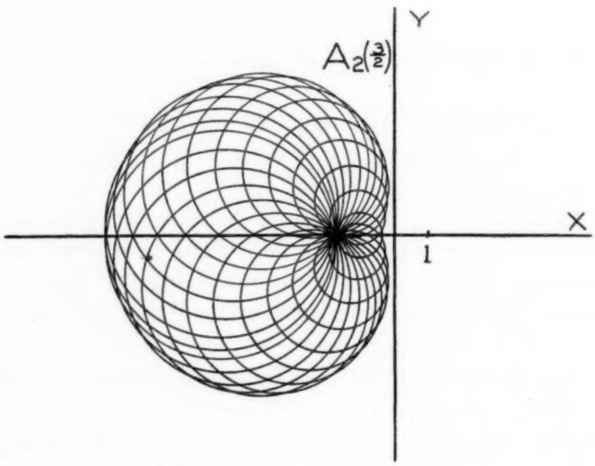

FIG. 1. The Region $A_2(\frac{3}{2})$

absolute values are obtained respectively by considering products of the type z_1c_1 and z_2c_2.

Let us set $k = h(h^2 - 1)$ and $a = h^2/(h^2 - 1)$. Then $z_1 = d(\alpha)e^{i\alpha}$ and $c_1 = d(\beta)e^{i(\beta+\pi)}$, where

$$d(\theta) = a \cos \theta - (k^2 - a^2 \sin^2 \theta)^{\frac{1}{2}}.$$

It follows from considerations of elementary calculus that for $\alpha + \beta = \gamma$ fixed, the expression $|z_1c_1|$ is a minimum when $\alpha = \beta = \gamma/2$. We note that $\gamma/2$ cannot exceed the value arc sin $1/h$.

Substituting the values $\alpha = \beta = \gamma/2 = (\theta - \pi)/2$ and $a = h^2/(h^2 - 1)$, $k = h/(h^2 - 1)$ we then have as part of the boundary of the region $A_2(h)$

$$r = d^2(\gamma/2) = \left(\frac{h}{h^2 - 1}\right)^2 [1 - h^2 \cos \theta - 2h \sin \theta/2(1 - h^2 \cos^2 \theta/2)^{1/2}]$$

where $\pi - 2\lambda \leqq \theta \leqq \pi + 2\lambda$; $\lambda = \text{arc} \sin 1/h$.

An analogous argument shows that the remaining portion of the boundary is given by the equation

$$r = \left(\frac{h}{h^2 - 1}\right)^2 [1 - h^2 \cos\theta + 2h \sin\theta/2(1 - h^2 \cos^2\theta/2)^{1/2}],$$

$$(\pi - 2\lambda \leqq \theta \leqq \pi + 2\lambda).$$

It follows that $A_2(h)$ is the region outside the curve

$$r^2 + 2r(a^2 \cos\theta - k^2) + a^2 = 0, \tag{12}$$

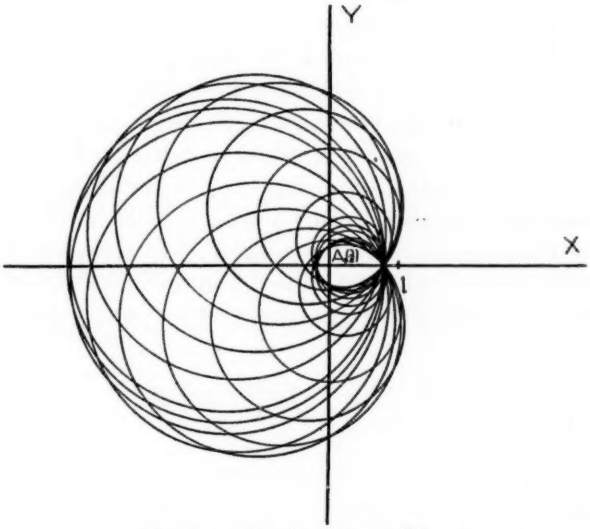

FIG. 2. The Region $A_1(\frac{3}{2})$

where we recall that

$$a = \frac{h^2}{h^2 - 1}, \qquad k = \frac{a}{h},$$

and h is any number > 1.

Similarly, the condition that $a_{2n+1} = re^{i\theta}$ be an element of the region $A_1(h)$ can be shown to be

$$r < \frac{1}{(h^2 - 1)^2}[h^2 - \cos\theta - 2\sin\theta/2(h^2 - \cos^2\theta/2)^{1/2}]. \tag{13}$$

The following theorem has now been proved.

THEOREM 3. *The continued fraction*

$$\frac{1}{1} + \frac{a_2}{1} + \frac{a_3}{1} + \cdots$$

converges if the numbers $a_{2n} \epsilon A_2(h)$ *[formula* (12)] *and the numbers* $a_{2n+1} \epsilon A_1(h)$ *[formula* (13)], *for some* $h > 1$, *and if in addition the numbers* a_n *are bounded away from the boundaries of these regions.*

It is of course evident that the roles of the odd and of the even a's can be interchanged.

As $h \rightarrow 1$ the regions $A_2(h)$ and $A_1(h)$ both tend to the parabola found by Scott and Wall [2].

We now note that if the parameter h is taken large enough, the region $A_2(h)$ can be made to include all the plane except an arbitrary small neighborhood of the point $z = -1$. The region $A_1(h)$ is then restricted to a small neighborhood of the point $z = 0$. The following is a corollary of Theorem 3.

COROLLARY. *The continued fraction* (1) *converges at least in the wider sense, if* $\lim a_{2n+1} = 0$ *and if* $z = -1$ *is not a limit point of the sequence* $\{a_{2n}\}$.

Under the conditions of the corollary there exists an $e > 0$ and a positive integer $n(e)$ such that $|1 + a_{2n}| > e$, when $n \geqq n(e)$. Further, there exists a number h such that $a_{2n} \epsilon A_2(h)$ for all $n \geqq n(e)$. Finally we can find a positive integer $n(h)$ such that for all $n \geqq n(h)$, $a_{2n+1} \epsilon A_1(3h)$. If we set $N = \max [n(e), n(h)]$, the numbers a_{2n} and $a_{2n+1} (n \geqq N)$ will certainly lie in the regions $A_2(2h)$ and $A_1(2h)$ respectively and will be bounded away from the boundaries of these regions. Theorem 3 therefore insures the convergence of the continued fraction

$$\frac{1}{1} + \frac{a_{2N}}{1} + \frac{a_{2N+1}}{1} + \cdots .$$

From a well-known argument it follows that the original continued fraction converges at least in the wider sense. The proof of the corollary is complete.

This result permits the following application to continued fractions of the form (2).

THEOREM 4. *Let* L *(possibly including the point* ∞) *be the set of limit points of the sequence* $\{a_{2n}\}$ *of the continued fraction*

$$1 + \frac{a_1 x}{1} + \frac{a_2 x}{1} + \cdots ,$$

and suppose $\lim a_{2n+1} = 0$. *The continued fraction then converges to a meromorphic function of the complex variable* x *in every region contained in the set* D, *where* D *is defined as follows:* $x \epsilon D$ *if*

$$|x| < M,$$

and if

$$\left| x + \frac{1}{k} \right| > e$$

for all $k \epsilon L$, $k \neq 0$. *The constants* M *and* e *are any positive numbers such that* $e \leqq 4M$.

Consider the set of numbers a_{2n}. They may be distributed into three categories as follows. If the point at infinity belongs to L, the first category A will contain all a_{2n} such that $| a_{2n} | \geqq 2/e$. If infinity does not belong to L, A is to be the null set. By hypothesis $| x | > e$ for all x in D and hence $| 1 + a_{2n}x | > 1 > e/4M$ for $a_{2n} \epsilon A$. The second category B contains all a_{2n} such that $| a_{2n} | < 1/2M$. This set may be null. Each member of B has the property that $| a_{2n}x | < 1/2$ and hence that $| 1 + a_{2n}x | > 1/2 > e/4M$, for all x in D. Finally, the third category C contains the remaining numbers a_{2n}. Except for at most a finite number of elements of C, the members of C and hence the limit points of C are in modulus $\geqq 1/2M$ and $\leqq b$, where b is a suitably chosen finite constant $\geqq 2/e$.

By Zermelo's principle, if needed, numbers k_n belonging to L and a positive integer n_0 can be chosen such that for $n \geqq n_0$

$$\left| \frac{a_{2n}}{k_n} - 1 \right| < \frac{e}{4M} \qquad (a_{2n} \epsilon C).$$

It follows from our hypothesis on D that

$$\left| a_{2n} x + \frac{a_{2n}}{k_n} \right| > \frac{e}{2M}$$

for all x in D. Combining these two relations we see that

$$| 1 + a_{2n} x | \geqq \left| a_{2n} x + \frac{a_{2n}}{k_n} \right| - \left| \frac{a_{2n}}{k_n} - 1 \right| > \frac{e}{4M},$$

for a_{2n} in $C (n \geqq n_0)$ and all x in D.

A number h independent of x thus exists such that $a_{2n}x \epsilon A_2(h)$, where for all x in D and all $n \geqq n_0$ the quantities $a_{2n}x$ are bounded away from the boundary of $A_2(h)$. Further, since $\lim a_{2n+1} = 0$ and $| x | < M$, a positive integer n_1 exists such that for all $n \geqq n_1$ and all x in D, $a_{2n+1}x \epsilon A_1(h)$ where the set of all $a_{2n+1}x$ is bounded away from the boundary of $A_1(h)$. If we now let $N = \max(n_0, n_1)$, it follows that the continued fraction

$$\frac{1}{1} + \frac{a_{2N} x}{1} + \frac{a_{2N+1} x}{1} + \cdots \tag{14}$$

converges. The uniform convergence of this continued fraction follows from Theorem 2. The quantities z mentioned there are rational functions of the variable x,

$$z_n = 1 + \frac{a_n x}{1} + \frac{a_{n-1} x}{1} + \cdots + \frac{a_2 x}{1},$$

and have no pole for x in D as for all x in D the system of equations (10) has unique and finite solutions. The continued fraction (14) therefore converges to a regular analytic function in every region contained in D. A well-known argument then insures the convergence of the complete continued fraction to a

function meromorphic in every region contained in D. This completes the proof of Theorem 4.

To illustrate what may happen when $\lim a_{2n} = -1$ and $\lim a_{2n+1} = 0$, we give two examples. The continued fraction (1) diverges if $a_2 = -2$, $a_{2n} = -(n+1)/(n-1)$ and $a_{2n+1} = -1/n^2$; it converges if $a_2 = -2$, $a_{2n} = -(n+1)(n^2-3n+3)/(n-1)^3$ and $a_{2n+1} = -1/n^4$. These results can be quickly verified by a reference to equations (10).

5. Value regions

As a by-product of the method used to establish the above convergence criteria we obtain results on value regions. The connection is established by the following lemma.

LEMMA. *If there exist in the complex plane four regions A_1, A_2, V_1, V_2 such that*

$$\text{i)}\quad \begin{aligned}&\text{(a) } 1 + A_1 \subset V_1,\\ &\text{(b) } 1 + A_2 \subset V_2,\end{aligned}$$

$$\text{ii)}\quad \begin{aligned}&\text{(a) } 1 + a/v \in V_1, \text{ if } a \in A_1,\ v \in V_2,\\ &\text{(b) } 1 + a/v \in V_2, \text{ if } a \in A_2,\ v \in V_1,\end{aligned}$$

then all approximants of the continued fraction

$$1 + \frac{a_1}{1} + \frac{a_2}{1} + \cdots \tag{15}$$

lie in the region V_1, and all approximants of

$$1 + \frac{a_2}{1} + \frac{a_3}{1} + \cdots \tag{16}$$

lie in V_2, if $a_{2n} \in A_2$, $a_{2n+1} \in A_1$.

Let

$$\frac{A_{2n}}{B_{2n}} = 1 + \frac{a_1}{1} + \cdots + \frac{a_{2n}}{1}.$$

It follows from condition (b) of i) that $1 + a_{2n} \in V_2$. Alternate application of conditions ii) (a) and ii) (b) then gives the desired result. The proof is analogous for the odd approximants of the first and all approximants of the second continued fraction.

We now note that the conditions of the lemma are satisfied for

$$A_1(h) = A_1,\ A_2(h) = A_2,$$

$$Z_1 = V_1,\ Z_2 = V_2.$$

The following theorem is then an immediate consequence of our previous results and the lemma.

THEOREM 5. *If* $h > 1$, $a_{2n} \epsilon A_2(h)$ *[formula* (12)*]*, $a_{2n+1} \epsilon A_1(h)$ *[formula* (13)*], then all approximants of the continued fraction* (15) *lie in the region* Z_1 *[formula* (11), (i)*] and all approximants of* (16) *lie in* Z_2 *[formula* (11), (ii)*].*

THE RICE INSTITUTE.

BIBLIOGRAPHY

1. O. PERRON, *Die Lehre von den Kettenbrüchen*, Leipzig, 1929.
2. W. T. SCOTT AND H. S. WALL, *A convergence theorem for continued fractions*, Trans. A. M. S. vol. 47 (1940) pp. 155–172.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

ON THE CONTINUUM HYPOTHESIS

By Isaiah Maximoff

(Received March 21, 1942)

In an earlier paper[1] the author proved that the set $E_x^{(r)}$ of all the sequences

$$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}, \alpha \leqq \beta < \Omega_r, \tag{1}$$

$$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}, \alpha < \Omega_r, \tag{2}$$

where x_α is any one of the numbers

$$1, 2, 3, \cdots \gamma, \cdots, \gamma < \Omega_r, \tag{3}$$

is a *continuum* of power $2^{\aleph_i}$.

Let (a, b) be a pair of elements of the set $E_x^{(r)}$ such that $a < b$.

Definition 1. *The set of all the elements x such that $a < x < b$ will be called an interval (a, b) with the boundaries a, b.*

Let E be a set of points in $E_x^{(r)}$.

Definition 2. *A point x_0 of $E_x^{(r)}$ will be called a limit $(\mathfrak{T})$-point for E if every interval containing x_0 contains at least one point x_1 of E different from x_0, $x_1 \neq x_0$.*

Let $E_{\mathfrak{T}}^{(1)}$ be the set of all the limit $(\mathfrak{T})$-points for E.

Definition 3. *We shall say that a set E is perfect $(\mathfrak{T})$ if $E_{\mathfrak{T}}^{(1)} = E$.*

Let $y = f(x)$ be a single valued function defined in $E \subset E_x^{(r)}$. Denote by $f(E)$ the set of all the points $y = f(x)$ where $x \in E$.

Definition 4. *The function $y = f(x)$ will be called regular in E if for every pair (x_1, x_2) of points x_1, x_2 of E such that $x_1 \neq x_2$ we have $f(x_1) \neq f(x_2)$.*

Definition 5. *The function $y = f(x)$ will be called continuous $(\mathfrak{T})$ in E if for every pair (x_0, y_0) where $y_0 = f(x_0)$, $x_0 \in E$, to every interval $i(y_0)$ containing y_0 there corresponds an interval $i(x_0)$ containing x_0 and having the following property: from $x_0 \in i(x_0)E$ there follows $y \in i(y_0)f(E)$. We then write: $i(y_0) \to i(x_0)$.*

Definition 6. *The set $E \subset E_x^{(r)}$ will be called hypermeasurable $(\mathfrak{T})$ if E either contains a perfect $(\mathfrak{T})$ set or is of the power $\leqq \aleph_r$.*

Definition 7. *A function $y = f(x)$ defined in E, will be called hypermeasurable $(\mathfrak{T})$ in E, if the set E contains a perfect $(\mathfrak{T})$ set $\mathfrak{P}$ in which $f(x)$ is continuous $(\mathfrak{T})$.*

Definition 8. *Let*

$$x = \{x_0, x_1, x_2, \cdots, x_\alpha, \cdots\}, \alpha < \Omega_{r+k}, \tag{4}$$

be a sequence of points x_α of the space $E_x^{(r)}$. This sequence will be called a point of the transfinite space E_x^{rk}.

Evidently, the power of the set E_x^{rk} is equal to $2^{\aleph_{r+k}}$. We shall say that a point

$$x' = \{x'_0, x'_1, x'_2, \cdots x'_\alpha \cdots\}, \alpha < \Omega_{r+k}, \tag{5}$$

of E_x^{rk} is equal to a point

[1] *On a Continuum of Power $2^{\aleph_i}$* (Annals of Mathematics, vol. 41, no. 2, April 1940).

$$(6) \qquad x'' = \{x_0'', x_1'', x_2'', \cdots x_\alpha'', \cdots\}, \alpha < \Omega_{r+k},$$

if $x_\alpha' = x_\alpha''$ for every $\alpha < \Omega_{r+k}$. In the contrary case we write $x' \neq x''$. We shall denote by $\rho\, I$ the point

$$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

of E_x^{rk} such that $x_0 = x_1 = x_2 = \cdots = x_\alpha = \cdots = \rho \in E_x^{(r)}$.

Let E be any set of points in E_x^{rk} and let

$$\bar{x} = \{\bar{x}_0, \bar{x}_1, \bar{x}_2, \cdots \bar{x}_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

be any point of E. We shall denote
(i) by $\{\bar{x}\}_\alpha$ the point $\bar{x}_\alpha$ of $E_x^{(r)}$, $\bar{x}_\alpha = \{\bar{x}\}_\alpha$;
(ii) by $\{E\}_\alpha$ the set of all the points $\{\bar{x}\}_\alpha$ where $\bar{x} \in E$. A set $\mathcal{T} \subset E_x^{rk}$ will be called an interval in E_x^{rk} if $\{\mathcal{T}\}_\infty$ is an interval in $E_x^{(r)}$ for every $\alpha < \Omega_{r+k}$. Furthermore we shall say that a set E, $E \subset E_x^{rk}$, is perfect ($\mathcal{T}$), hypermeasurable ($\mathcal{T}$) if the set $\{E\}_\alpha$ is perfect ($\mathcal{T}$), hypermeasurable ($\mathcal{T}$) respectively.

If we now suppose that to every point

$$\bar{x} = \{\bar{x}_0, \bar{x}_1, \bar{x}_2, \cdots \bar{x}_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

of a set $E \subset E_x^{rk}$ there corresponds one and only one point

$$y = f(\bar{x}) = \{f_0(\bar{x}_0), f_1(\bar{x}_1), f_2(\bar{x}_2), \cdots f_\alpha(\bar{x}_\alpha), \cdots\}, \alpha < \Omega_{r+k},$$

of E_x^{rk}, then we have a single valued function $y = f(x)$ defined in $E \subset E_x^{rk}$.

Denote by $f(E)$ the set of all the points $y = f(x)$ where $x \in E$. If $f(x_1) \neq f(x_2)$ in all cases when $x_1 \neq x_2$, $x_1 \in E$, $x_2 \in E$, then the function $f(x)$ is said to be *regular* in E.

A function $f(x)$ defined in E, $E \subset E_x^{rk}$, will be called continuous ($\mathcal{T}$), hypermeasurable ($\mathcal{T}$) in E if $\{f(x)\}_\alpha$ is respectively continuous ($\mathcal{T}$), hypermeasurable ($\mathcal{T}$) in $\{E\}_\alpha$ for every $\alpha < \Omega_{r+k}$. From this it follows that a function defined in $E \subset E_x^{rk}$ is hypermeasurable ($\mathcal{T}$) in E if E contains a perfect ($\mathcal{T}$) set $\mathcal{P}$ such that $f(x)$ is continuous ($\mathcal{T}$) in $\mathcal{P}$.

We denote by E_1 the set of all points

$$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}, \alpha < \Omega_{r+k},$$

of E_x^{rk} having the following property: any one of the elements $x_0, x_1, x_2, \cdots x_\alpha, \cdots$, $\alpha < \Omega_{r+k}$, of $E_k^{(r)}$ is the rational point $\{2\}$ of $E_x^{(r)}$ and every other element is the rational point $\{1\}$ of $E_x^{(r)}$.

Let $|\bar{\alpha}|$ be the point

$$x = \{x_0, x_1, x_2, \cdots x_\alpha, \cdots\}$$

of E_x^{rk} such that $x_\alpha = \{2\}$ and $x_\alpha = \{1\}$ for $\alpha_1 \neq \alpha$.

We denote by E_2 the set of all points $\rho\, I$ where $\rho \in E_x^{(r)}$ and $I = \{1, 1, 1, \cdots 1, \cdots\} \in E_x^{rk}$.

Now we consider a correspondence

$$\alpha = Z(\rho) \tag{7}$$

between the number α, $1 \leqq \alpha < \Omega_{r+k}$, and the point $\rho \in E_x^{(r)}$. It is easily seen that we can replace the correspondence (7) by the equivalent correspondence

$$|\underline{\alpha}| = Z(\rho I) \tag{8}$$

between the points

$$|\underline{\alpha}|, \rho I$$

of E_x^{rk}.

We shall say that $f(x)$ is a function of Nicholas Parfentieff if

(i) $\qquad f(x)$ is regular in E_2;

(ii) $\qquad E_1 \neq f(E_2)$.

We denote by (P) the class of all the functions of Nicholas Parfentieff.

Theorem 1. *If a function $f(x)$ is continuous $(\mathcal{T})$ and regular in $E_2 \subset E_x^{rk}$, then $f(x)$ is a function of N. Parfentieff.*

Proof. Suppose the function $f(x)$ does not belong to the class (P). Since $f(x)$ is regular in E_2, we have: $E_1 = f(E_2)$. Since the function $y = f(x)$ is continuous $(\mathcal{T})$ in E_2, to an interval $\mathcal{T}(y_0) \subset E_x^{rk}$ $(y = f(x_0), x_0 \in E)$ there corresponds an interval $\mathcal{T}(x_0)$ in such a manner that when x runs over the set $\mathcal{T}(x_0)E_2$ then $f(x)$ runs over a set H contained in $\mathcal{T}(y_0)\, f(E_2)$. We then write: $\mathcal{T}(y_0) \to \mathcal{T}(x_0)$. Let $\mathcal{T}_\nu(y)$ be an interval in E_x^{rk} containing the point $|\underline{\nu}| \in E_x^{rk}$. Evidently, to the interval $\mathcal{T}_\nu(y)$ there corresponds an interval $\mathcal{T}(x)$, $\mathcal{T}_\nu(y) \to \mathcal{T}(x)$, where $y = f(x)$. It is obvious that the set $\mathcal{T}(x)E_2$ has the power $\geqq \aleph_0$, but the set $\mathcal{T}_\nu(y)E_1$ consists of one and only one point $|\underline{\nu}|$. Since the function $f(x)$ is regular in E_2, the set H of all the points $f(x)$ where $x \in \mathcal{T}(x)E_2$ also has the power $\geqq \aleph_0$. However this is impossible in view of $H \subset \mathcal{T}_\nu(y)E_1$.

Theorem 2. *Every hypermeasurable $(\mathcal{T})$ regular in E_2 function $y = f(x)$ is a function of N. Parfentieff.*

Proof. Suppose the function $y = f(x)$ does not belong to the class (P). Since $f(x)$ is regular in E_2, then $E_1 = f(E_2)$. Since the function $f(x)$ is hypermeasurable $(\mathcal{T})$ in E_2, there exists a perfect $(\mathcal{T})$ set F contained in E_2 and such that $f(x)$ is continuous $(\mathcal{T})$ in F. Let $E_1' = f(F)$ and let $|\underline{\mu}|$ be any point of E_1'. It is clear that to an interval $\mathcal{T}_\mu(y)$, $\mathcal{T}_\mu(y) \subset E_x^{rk}$, $(y = f(x), x \in F)$ there corresponds an interval $\mathcal{T}(x) \subset E_x^{rk}$ in such a manner that when x runs over the set $\mathcal{T}(x)F$, then $y = f(x)$ runs over a set H' contained in $\mathcal{T}_\mu(y)E_1'$; for $f(x)$ is regular in E_2 and since $\mathcal{T}_\mu(y)E_1'$ consists of one and only one point. But the set $\mathcal{T}(x)F$ has the power $\geqq \aleph_0$, and so H' has also the power $\geqq \aleph_0$. This is impossible since $H' \subset \mathcal{T}_\mu(y)E_1'$.

Thus, in the mathematics $\mathfrak{L}$ of the hypermeasurable $(\mathcal{T})$ functions we have: $2^{\aleph_r} \neq \aleph_{r+k}$ $(0 \leqq r < \Omega_0, 1 \leqq k < \Omega_0)$.

Tchebolssary, Tchapaewskiposelok, U.S.S.R.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

LINEARLY ARC-WISE CONNECTED TOPOLOGICAL ABELIAN GROUPS

By R. C. James

(Received December 10, 1941)

Several types of connectedness of topological spaces can be defined in terms of continuous functions with real numbers as arguments and values in the space.[1] The function can be required to be linear if the space is a topological abelian group, giving rise to properties that can be used as necessary and sufficient conditions for the topological abelian group to be a linear topological space. If the concepts of convexity and boundedness are generalized to topological abelian groups, these properties lead to a theorem on normability analogous to one given by Kolmogoroff for linear topological spaces.[2] I am indebted to Dr. Michal for suggesting an investigation of linear arc-wise connectedness, and for his guidance.

1. Topological groups and linear topological spaces

By a topological group, we mean an abstract group which has a Hausdorff topology[3] with respect to which the group operations are continuous. A topological abelian group is a topological group whose abstract group is abelian. A linear topological space is a linear space[4] which has a Hausdorff topology with respect to which the fundamental operations $x + y$ and ax are continuous. It can easily be shown that a linear topological space is a topological abelian group and has the following properties:

I. If $ax = 0$, either $x = 0$ or $a = 0$.
II. If $ax = ay$ and $a \neq 0$, then $x = y$; if $ax = bx$ and $x \neq 0$, then $a = b$.
III. If U is an open set and $a \neq 0$, then aU is an open set.
IV. If U is a neighborhood of the origin, there is a neighborhood V of the origin such that $aV \subset U$ for all a satisfying $|a| \leqq 1$.

2. Continuous functions and arc-wise connectedness

The functions treated in this paper are exclusively functions with real numbers as arguments and values in a topological abelian group or a linear topological space. Such a function $f(t)$ is said to be continuous at $t = t_1$ if for any neighborhood U of $f(t_1)$ there is a $\delta > 0$ such that $f(t) \in U$ for all t satisfying $|t - t_1| < \delta$. A continuous function will be called linear if it is additive [i.e. $f(t_1 + t_2) = f(t_1) + f(t_2)$] and continuous throughout its interval of definition. It will be convenient to introduce the concept of uniform continuity:

Definition 2.1. *A function $f(t)$ with real numbers as arguments and values in a topological group T is said to be uniformly continuous in an interval (a, b) if for any*

[1] See Pontrjagin (VI) for a discussion of these properties. (Roman numerals refer to the bibliography.)
[2] Kolmogoroff (IV).
[3] Hausdorff (II). The space satisfies (A), (B), (C), (5), pp. 228–229.
[4] Banach (1), pg. 26.

neighborhood U of the identity there is a positive number δ such that $f(t) \in f(t_1) + U$ if t_1 and t belong to (a, b) and $| t_1 - t | < \delta$.

This definition can be readily extended to functions of several variables. It can be shown that a uniformly continuous function is continuous and that a function which is continuous in a closed interval is uniformly continuous in this interval.

Several types of arc-wise connectedness of topological spaces can be defined in terms of continuous functions of a real variable, two points x and y of the topological space being said to be joined by a curve or arc if there is a continuous function $f(t)$ such that $f(0) = x$ and $f(1) = y$.[5] Only one of these types of connectedness is needed in this paper (def. 2.2), although the others will be used for topological abelian groups in a more restricted sense as defined by means of linear functions.

DEFINITION 2.2. *A topological space is called simply connected if every closed path in the space can be continuously shrunk to a point; explicitly, if for every closed path $f(t)$ for which $f(0) = f(1)$ there is a function $g(s, t)$, continuous simultaneously in s and t, for which $g(0, t) \equiv f(t)$, $g(1, t) \equiv f(0)$, and $g(s, 0) \equiv g(s, 1) \equiv f(0)$.*

DEFINITION 2.3. *A topological abelian group T is called linearly arc-wise connected if: (1) the identity can be joined to any point of T by a curve defined by a linear function; or (2) any two points x_1 and x_2 of T can be joined by a curve defined by an affine function, i.e. a function $f(t)$ satisfying the functional relation: $f(0) + f(t_1 + t_2) = f(t_1) + f(t_2)$.[6] If there is only one linear path joining the identity to any given point of T, then T will be called uniquely linearly arc-wise connected.*

DEFINITION 2.4. *A topological abelian group T is called locally linearly arc-wise connected if: (1) for any neighborhood U of the identity there is a neighborhood V of the identity such that the identity can be joined to any point of V by a curve contained in U and defined by a linear function; or (2) for any neighborhood U' of a point x_1 of T there is a neighborhood V' of x_1 such that any point x_2 of V' can be joined to x_1 by a curve contained in U' and defined by an affine function.*

It can be shown that the two forms of definition 2.3 (or 2.4) are equivalent. From the fact that a connected topological group can be generated by an arbitrary neighborhood of the identity,[7] it follows that a topological abelian group which is connected and locally linearly arc-wise connected is also linearly arc-wise connected. A linear topological space can be shown to have the properties of definitions 2.2, 2.3, and 2.4. In fact, the linear path joining the identity to any point x is $f(t) \equiv tx$, and any closed path $f(t)$ can be continuously deformed to a point by means of the function $g(s, t) \equiv sf(0) + (1 - s)f(t)$. Property IV of section 1 shows that a linear topological space is locally linearly arc-wise connected.

[5] See Pontrjagin (VI), pp. 219–221, for explicit definitions.

[6] The definition of *curve* implies that $f(t)$ is also continuous.

[7] Pontrjagin (VI), pg. 76, theorem 15.

3. Points of finite order in a topological abelian group

Since a linear topological space has no non-zero points of finite order, this is a necessary condition for a topological abelian group to be a linear topological space. It will be shown in this section that a topological abelian group which is connected, simply connected and locally linearly arc-wise connected must have either no non-zero elements of finite order or the set of elements of the n^{th} order is dense in itself for each $n > 1$. However topological abelian groups can be found which are linearly arc-wise connected and locally linearly arc-wise connected and have any desired number of elements of a given prime order, provided an abstract abelian group exists which has this number of elements of that order. This is shown by the following examples:

Let T be composed of elements of the form $x = (x_1, x_2, \cdots), y = (y_1, y_2, \cdots)$, where x_i and y_i are any complex numbers of absolute value 1. Define $x + y$ as $(x_1y_1, x_2y_2, \cdots)$ and a neighborhood U of x as the set of elements of the form $u = (u_1, u_2, \cdots)$, where $| u_i - x_i | < \epsilon$ for $i \leqq 1/\epsilon$. It can be shown that T is a topological abelian group which is linearly arc-wise connected and locally linearly arc-wise connected (but not simply connected) and the set of elements of the n^{th} order is dense in itself for each $n > 1$. It is an open question whether this is possible if the space is also simply connected. Let T' be composed of elements of the form $x = (x_1, x_2, \cdots, x_r)$, where x_i is any complex number of absolute value 1, and addition and neighborhoods are defined as above. If p is a prime, the element x is of the p^{th} order if and only if $x_j = \exp(2\pi i(n_j/p))$ for each j, and at least one n_j is not a multiple of p. Hence T' contains $p^r - 1$ elements of the p^{th} order, and any abstract abelian group contains $p^r - 1$ elements of the p^{th} order (for some integer r).

The following three lemmas will be used in establishing the theorem of this section:

LEMMA 3.1. *If a topological abelian group is linearly arc-wise connected and has a non-zero element of finite order, then for each positive integer n there are at least $n - 1$ non-zero elements x_i such that $nx_i = 0$.*

LEMMA 3.2. *If a topological abelian group is locally linearly arc-wise connected, and for some integer $p > 1$ there is an element of order p in every neighborhood of the identity, then this is true for any positive integer n, and the set of elements of order n is dense in itself (for each integer $n > 1$).*

Lemma 3.1 and the first part of lemma 3.2 can be readily proved by exhibiting certain closed linear paths which are not identically zero. Connectedness is not necessary in Lemma 3.2, since a closed linear path in some neighborhood of the identity is sufficient. The proof that the set of elements of order n $(n > 1)$ is dense in itself if there is an element of order n in every neighborhood of the identity is as follows:

Let x be any element of order n and W a neighborhood of x. Then $W - x$ is an open set containing the identity. Let $U \subset W - x$ be a neighborhood of the identity which contains none of the elements ix $(i = 1, 2, \cdots, n - 1)$, and

V be a neighborhood of the identity such that $V^n \subset U$.[8] There is by assumption an element y of order n belonging to V. Then $x + y \in W$ and $n(x + y) = nx + ny = 0$. If $i(x + y) = 0$ for $0 < i < n$, then $iy = -ix = (n - i)x$. But this is impossible, since $y \in V$, $iy \in V^i \subset U$, and $(n - i)x$ does not belong to U. Hence for each positive integer n the set of elements of order n is dense in itself.

LEMMA 3.3. *If a topological abelian group is locally linearly arc-wise connected and there is a positive integer n and a neighborhood U of the identity such that $x \in U$ and $nx = 0$ imply $x = 0$, then for any neighborhood M of the identity there is a neighborhood $N \subset M$ such that if $y \in x + N$, and x_1 is any point such that $nx_1 = x$, there is a unique point y_1 which belongs to $x_1 + N$ and is such that $ny_1 = y$.*

PROOF: For such a neighborhood U and any neighborhood M of 0, take a neighborhood V such that $V \subset M$ and $V \subset U$. Let N be the set of all points of V which can be joined to the identity by a linear path $f(t)$ for which there is a neighborhood L of the identity such that $f(t) + L \subset V$. Suppose $nx_1 = x$ and $y \in x + N$, or $y - x \in N$. Let $g(t)$ be the linear path joining the identity to $y - x$. Then $y_1 = x_1 + g(1/n) \in x_1 + N$, and $ny_1 = y$. Since $N \subset U$, there is no other point y_2 of $x_1 + N$ such that $ny_2 = y$.

It now remains to show that N is open. Let x be a point of N, $f(t)$ the linear path joining the identity to x, L the neighborhood of the identity such that $f(t) + L \subset V$, and L' a neighborhood of the identity such that $2L' \subset L$. Since T is locally linearly arc-wise connected, there is a neighborhood L'' of the identity such that if $y \in L''$ there is a linear path contained in L' which joins the identity to y. Then $L'' + x$ is an open set contained in N, for if $z \in L'' + x$, then $z - x \in L''$, and hence there is a linear path $g(t)$ contained in L' and joining the identity to x. Then $[g(t) + f(t)] + L' \subset V$ and $g(t) + f(t)$ joins the identity to z. Hence $z \in N$.

THEOREM 3.1. *If a topological abelian group T is connected, simply connected and locally linearly arc-wise connected, then T either has no non-zero elements of finite order, or else for each positive integer $n > 1$ the set of elements of the nth order is dense in itself.*

PROOF: Suppose there is a non-zero element of finite order and the points x_i for which $2x_i = 0$ is not dense in itself; that is, there is a neighborhood U of the identity such that $x \in U$ and $2x = 0$ imply $x = 0$. Since T is connected and locally linearly arc-wise connected, it is linearly arc-wise connected. Let $f(t)$ be a linear function of t for which $f(0) = 0$, $f(1) = x \neq 0$, where $x \in T$. By lemma 3.1, there is an element y of order two. Then $2[f(1/2) + y] = x$, and since T is linearly arc-wise connected, there is a function $F(t)$ which is continuous for $0 \leqq t \leqq 1$ and such that $F(0) = f(1/2)$, $F(1) = f(1/2) + y$. Also, $2F(t)$ is a closed path beginning and ending at x, and hence there is a function $g(s, t)$, continuous simultaneously in s and t, such that $g(0, t) \equiv 2F(t)$, $g(1, t) \equiv x$, $g(s, 0) \equiv g(s, 1) \equiv x$.

Take a neighborhood $N \subset U$ of the identity, as defined in lemma 3.3. There

[8] U^n is the set of all $y = u_1 + u_2 + \cdots + u_n$, where each u_i belongs to U.

is an ϵ_1 such that if $0 \leqq s < \epsilon_1$, $| t' - t | < \epsilon_1$, and $0 \leqq t \leqq 1$, then $g(s, t) \in N + g(0, t')$ for all t' $(0 \leqq t' \leqq 1)$.[9] Let $0 < s_1 < \epsilon_1$, and hence $g(s_1, t') \in g(0, t') + N$ for all t'. Since N has the properties of lemma 3.3, there is a unique point $x_1 \in N + F(t')$ such that $2x_1 = g(s_1, t')$. Define the function $F(s_1, t)$ by the relations $F(s_1, t') = x_1$. The function $F(s_1, t)$ is uniquely defined for all t $(0 \leqq t \leqq 1)$; and $F(s_1, 0) = F(0)$, $F(s_1, 1) = F(1)$, $2F(s_1, t) \equiv g(s_1, t)$.

It will now be shown that $F(s_1, t)$ is continuous in t. Let U be any neighborhood of $F(s_1, t')$. Since $F(s_1, t') + 0 \subset F(t') + N$, there is a neighborhood U_1 of $F(s_1, t')$ and a neighborhood V of 0 such that $U_1 + V \subset F(t') + N$ and $U_1 \subset U$. If δ_1 is such that $F(t') - F(t) \subset V$ for $| t' - t | < \delta_1$, then $U_1 + F(t') - F(t) \subset F(t') + N$, or $U_1 \subset F(t) + N$, for $| t' - t | < \delta_1$. Similarly, there is a δ_2 and a neighborhood U_2 of $g(s_1, t')$ such that $U_2 \subset g(0, t) + N$ if $| t' - t | < \delta_2$. Since $U_1 - F(s_1, t')$ and $U_2 - g(s_1, t')$ are both neighborhoods of the identity, there is a neighborhood M of the identity which is contained in their intersection. Take $N' \subset M$ and having the property of Lemma 3.3. Let δ be a positive number equal to or less than the smaller of δ_1 and δ_2, and such that if $| t' - t | < \delta$, then $g(s_1, t) \in g(s_1, t') + N'$. If $| t' - t | < \delta$, then there is a unique point x_1 in $F(s_1, t') + N'$ such that $2x_1 = g(s_1, t)$. But since $F(s_1, t)$ is the unique point in $F(t) + N$ such that $2F(s_1, t) = g(s_1, t)$, and $x_1 \in F(s_1, t') + N' \subset U_1 \subset F(t) + N$, it follows that $x_1 = F(s_1, t)$. That is, if $| t' - t | < \delta$, then $F(s_1, t) \in F(s_1, t') + N' \subset U_1 \subset U$. Hence $F(s_1, t)$ is a continuous function of t.

There is an ϵ (the same for all s_1, t_1) such that $g(s, t) \in g(s_1, t_1) + N$ if $| s_1 - s | < \epsilon$, $| t_1 - t | < \epsilon$, and s and t are in the interval (0, 1). Hence the above process can be continued by "jumps" of ϵ. That is, if $F(s_1, t)$ is a continuous function of t and $F(s_1, 0) = F(0)$, $F(s_1, 1) = F(1)$, $2F(s_1, t) \equiv g(s_1, t)$; then a continuous function of t, $F(s, t)$, can be defined for any s $(s_1 < s < s_1 + \epsilon$, $0 \leqq s \leqq 1)$, such that $F(s, 0) = F(0)$, $F(s, 1) = F(1)$, and $2F(s, t) \equiv g(s, t)$. But $F(1, t)$ is a continuous function of t and $F(1, 0) = F(0)$, $F(1, 1) = F(1)$, and $2F(1, t) \equiv g(1, t) \equiv x$. This contradicts the fact that in each of the neighborhoods $F(0) + N$ and $F(1) + N$ of $F(0)$ and $F(1)$, respectively, there is only one point x_1 such that $2x_1 = x$. That is, the assumption that there is a point of finite order and a neighborhood U of the identity such that $x \in U$ and $2x = 0$ imply $x = 0$ has been shown to be impossible. Hence if there is a non-zero point of finite order, there is a point of the second order in every neighborhood of the identity. Then from Lemma 3.2 it follows that the set of points of order n is dense in itself (for each integer $n > 1$).

4. Conditions for a topological abelian group to be a linear topological space

The following lemma will be needed to establish the results of this section. It is equivalent to the continuity of multiplication by real numbers, as defined in the proof of theorem 4.1.

LEMMA 4.1. *If a topological abelian group T is uniquely linearly arc-wise*

[9] See definition 2.1 and the following discussion.

connected and locally linearly arc-wise connected, then the function $F(t, x) = f(t)$, where $f(t)$ defines the linear path joining the identity to x, is simultaneously continuous in t and x.

PROOF: Let x' and t' be any values of x and t, and W a neighborhood of $f(t') = F(t', x')$, where $f(t)$ is the linear path joining the identity to x. Since $0 + f(t') = f(t')$, there are neighborhoods V_1 and V_2 of the identity and $f(t')$, respectively, such that $V_1 + V_2 \subset W$. Since T is locally linearly arc-wise connected, there is a neighborhood V' of the identity such that each point of V' can be joined to the identity by a linear path contained in V_1. Take $U = x' + V'$ and $x \epsilon U$. Let $g(t)$ and $f(t)$ be the linear paths joining the identity to x and x', respectively. Then $g(t) - f(t)$ is the linear path joining the identity to $x - x' \epsilon V'$, and is therefore contained in V_1. Since $f(t)$ is continuous, there is an $\epsilon > 0$ such that $f(t) \epsilon V_2$ for all t such that $| t' - t | < \epsilon$. If $| t' - t | < \epsilon$, then $g(t) - f(t) + f(t) = g(t)$ belongs to W. Since $g(t)$ was the linear path joining the identity to an arbitrary point x of U, it then follows that $F(t, x)$ is simultaneously continuous in t and x.

THEOREM 4.1. *A necessary and sufficient condition that a topological abelian group T be a linear topological space is that T be uniquely linearly arc-wise connected and locally linearly arc-wise connected.*

PROOF: For each real number a and point x of T, define $a \cdot x$ as $F(a, x) = f(a)$, where $f(t)$ is the linear function joining the identity to x.[10] Lemma 4.1 states that this multiplication is continuous simultaneously in a and x. From the fact that $f(t)$ is additive, it also follows that the multiplication satisfies the postulates for a linear space. Hence T is a linear topological space.

Several variations of this theorem can be gotten by giving other conditions in place of the unique linear arc-wise connectedness. It can be shown that if the linear path joining the identity to some point x is unique, or if there are no non-zero points of finite order, then the topological abelian group is uniquely linearly arc-wise connected if it is linearly arc-wise connected.

COROLLARY. *A necessary and sufficient condition that a topological abelian group T be a linear topological space is that T be connected, locally linearly arc-wise connected, and for some element $x \epsilon T$ there is only one linear path joining the identity to x.*

COROLLARY. *A necessary and sufficient condition that a topological abelian group be a linear topological space is that it be connected, locally linearly arc-wise connected, and possess no non-zero elements of finite order.*

If a connected topological abelian group T is locally linearly arc-wise connected, it is also linearly arc-wise connected. If T is also simply connected, it then follows from theorem 3.1 that T either contains no non-zero elements of finite order, or the set of elements of order n is dense in itself for each $n > 1$. Hence the second corollary above can be written in the form:

THEOREM 4.2. *A necessary and sufficient condition that a topological abelian*

---

[10] $f(t)$ is only required to be defined for $0 \leqq t \leqq 1$, but it can be uniquely extended to all real numbers by using the additive property.

group be a linear topological space is that it be connected, simply connected, locally linearly arc-wise connected, and there exist an integer $n > 1$ *and a neighborhood* U *of the identity such that* U *contains no elements of order* n.

5. Convex topological groups

A. Kolmogoroff and A. Tychonoff have called a neighborhood U of the origin of a linear topological space convex if $ax + (1 - a)y$ is in U for any elements x and y of U and real number a in the interval $(0, 1)$.[11] A linear topological space is said to be *locally convex* if for any neighborhood U of the origin there is a convex neighborhood V contained in U. John von Neumann calls a linear topological space *convex* if for any neighborhood U of the origin there is a neighborhood $V \subset U$ such that $2V = V^2$.[12] J. V. Wehausen has shown that these concepts of convexity and local convexity are equivalent.[13] Convexity can be extended to topological groups in a number of ways, the following seeming to be the most satisfactory.

DEFINITION 5.1. *A topological group is called convex if for every neighborhood* U *of the identity there is a neighborhood* $V \subset U$ *such that* $nx \in V^n$ *implies* $x \in V$ *(for every positive integer* n*).*

When the topological group is a linear topological space, this definition is equivalent to the above definitions of convexity and local convexity. It is also evident that a convex topological group has no non-zero points of finite order. The definition of convexity given by von Neumann can be applied to topological groups without rewording, but is not entirely satisfactory, since for each y in some neighborhood of the identity it implies the existence of an x such that $2x = y$. Convexity as given in definition 5.1 also has the advantage of being a consequence of normability (def. 6.4).

LEMMA 5.1. *A topological abelian group* T *which is convex and linearly arc-wise connected is locally linearly arc-wise connected.*

PROOF: Take any neighborhood U of the identity. Choose a neighborhood V such that $\bar{V} \subset U$ and $nx \in V^n$ implies $x \in V$. Let x be any point of V and $f(t)$ the linear path joining the identity to $x: f(0) = 0, f(1) = x$. Let p/q be any rational number in the interval $(0, 1)$, where p and q are positive integers. Then $qf(p/q) = f(p) \in V^p \subset V^q$. Hence $f(p/q) \in V$. Since $f(t)$ is continuous, it then follows that $f(t) \in \bar{V}$ for $0 \leqq t \leqq 1$. That is, each point of V can be joined to the identity by a linear arc contained in U, and hence T is locally linearly arc-wise connected.

THEOREM 5.1. *A necessary and sufficient condition that a topological abelian group be a convex linear topological space is that it be convex and linearly arc-wise connected.*

This theorem is an immediate consequence of the above lemma and theorem 4.1.

---

[11] Kolmogoroff (IV) and Tychonoff (VII).

[12] Neumann (V), pg. 4. $2V$ is the set of all $u = v + v$, where $v \in V$; $V^2 = V + V$ is the set of all $v_1 + v_2$, where v_1 and v_2 belong to V.

[13] Wehausen (VIII), pg. 158.

6. Bounded sets and normable topological abelian groups

Several equivalent definitions of bounded sets in linear topological spaces have been given. The following one is due to J. v. Neumann, and can be extended to topological groups. This property will be used in finding conditions for normability of topological abelian groups.

DEFINITION 6.1. *A set S of a linear topological space is called bounded if for any neighborhood U of the origin there is a number a such that $S \subset aU$.*[14]

DEFINITION 6.2. *A set S of a topological group will be called bounded if for any neighborhood U of the identity there exists an integer n such that $(1/m)S \subset U$ for all integers $m \geqq n$, where $(1/m)S$ is the set of all x such that $mx \in S$.*

DEFINITION 6.3. *A topological group will be called locally bounded if it has a bounded neighborhood of the identity.*

It can easily be shown that definition 6.2 is equivalent to definition 6.1 when the topological group is a linear topological space. Definition 6.1 could have been revised for topological groups by merely requiring that a be an integer. However this has the disadvantage of implying that for each element x of a bounded set $S \subset aU$ (where U is a neighborhood of the identity) of a topological group there exists an element u of U such that $x = au$. For example, if this definition were used, no non-zero elements of the multiplicative group of rational numbers would be bounded. Boundedness as given in definition 6.2 also has the advantage that local boundedness is implied by normability (def. 6.4), while the above example shows that this would not be true for the other form.

Even if the topological group were linearly arc-wise connected, def. 6.2 would not be equivalent to the form gotten by replacing "$(1/m)S \subset U$" by "$S \subset mU$". This is illustrated by the multiplicative group of all complex numbers with absolute value 1, for any single element is bounded in the revised sense, but no element is bounded in the sense of def. 6.2. This difference is closely related to the fact that the form "$(1/m)S \subset U$" implies that the set of all nx (for n an integer) is unbounded (symbolically, $\lim_{n\to\infty} nx = \infty$) if $x \neq 0$, while the second does not. Requiring that the relation $(1/m)S \subset U$ hold for all $m \geqq n$, rather than merely $(1/n)S \subset U$, has easily seen advantages, but becomes ambiguous when the topological group is a linear topological space.

DEFINITION 6.4. *A topological abelian group T is normable if to each point x of T there can be associated a non-negative real number $|x|$ which satisfies the conditions:*

(1). $|nx| = |n||x|$ *for all integers n and elements x of T.*

(2). $|x + y| \leqq |x| + |y|$.

(3). *The system of neighborhoods of T is topologically equivalent to the system of neighborhoods defined in terms of the distance between two points, as given by $\rho(x, y) = |x - y|$.*

THEOREM 6.1. *A necessary and sufficient condition that a linearly arc-wise*

[14] This and three other equivalent definitions are given in Hyers (III).

connected topological abelian group T be normable is that T be a normable linear topological space.[15]

PROOF: It is evident that the conditions of the theorem are sufficient. Now suppose that T is a normed linearly arc-wise connected topological abelian group. It follows from (1) and (3) of definition 6.4 that $|x| = 0$ if and only if $x = 0$. T then contains no non-zero points of finite order, for if $nx = 0$ it follows that $|nx| = 0 = |n||x|$, and hence that $x = 0$. It will now be shown that T is also locally linearly arc-wise connected. Let U be the neighborhood of the identity consisting of all points x for which $|x| < \epsilon_1$, and V be the neighborhood of the identity consisting of all points x for which $|x| < \epsilon_2$, where $\epsilon_2 < \epsilon_1$. Let y be a point of V and $f(t)$ the linear path joining the identity to y $[f(0) = 0, f(1) = y]$. Suppose $f(t)$ does not belong to V for some rational number $t_1 = p/q$, $0 < p/q < 1$, where p and q are positive integers. Then $|f(p/q)| \geqq \epsilon_2$, and $|f(p)| = q|f(p/q)| \geqq q\epsilon_2$. But since $|f(1)| < \epsilon_2$, it follows that $|f(p)| < p\epsilon_2$. These statements are contradictory, since $0 < p/q < 1$, or $q > p$. Hence $f(t) \epsilon V$ for all rational values of t in the interval $0 \leqq t \leqq 1$. From the continuity of $f(t)$ it follows that $f(t) \epsilon \bar{V} \subset U$ for all t. It has thus been shown that T is locally linearly arc-wise connected. T is then a linear topological space by the second corollary of theorem 4.1. It only remains to show that $|ax| = |a||x|$ for all real a. From definition 6.4, $|px| = |p||x|$ for all integers p. Hence $|(p/q)x| = |p||x/q| = |p/q||q||x/q| = |p/q||x|$. That is, $|ax| = |a||x|$ for all rational a. From the continuity of ax and the continuity of the norm, it now follows that $|ax| = |a||x|$ for all real a.

THEOREM 6.2. *A necessary and sufficient condition for the normability of a linearly arc-wise connected topological abelian group T is that T be convex and locally bounded.*

PROOF. If T is a normable, linearly arc-wise connected topological abelian group, then T is a normable linear topological space. If the system of neighborhoods is defined in terms of the norm, it is easily seen that each neighborhood is bounded and convex. Conversely, if the topological abelian group T is convex and linearly arc-wise connected, it follows from the second corollary of theorem 4.1 that T is a linear topological space. If T is also locally bounded, it follows from a theorem of Kolmogoroff's[16] that T is a normable linear topological space and hence a normable topological abelian group.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

BIBLIOGRAPHY

I. S. BANACH, *Théorie des Opérations Linéaires*, Warsaw (1932).
II. F. HAUSDORFF, *Mengenlehre*, Berlin (1927).

[15] A linear topological space is normable under the same conditions as given in definition 6.4, except that (1) is replaced by $|ax| = |a||x|$ for all real numbers a and points x of T. See Kolmogoroff (IV), pg. 30.

[16] Kolmogoroff (IV). "*A necessary and sufficient condition for the normability of a linear topological space is the existence of a bounded convex open set.*"

III. D. H. Hyers, *Locally bounded linear topological spaces*, Revista de Ciencias, vol. 41, pp. 555–574.
IV. A. Kolmogoroff, *Zur Normierbarkeit eines allgemeinen topologischen linearen Raumes*, Studia Mathematica, vol. 5, pp. 29–33.
V. John von Neumann, *On Complete Topological Spaces*, Transactions of the American Mathematical Society, vol. 37, pp. 1–20.
VI. I. L. Pontrjagin, *Topological Groups*, Princeton (1939).
VII. A. Tychonoff, *Ein Fixpunktsatz*, Math. Annalen, vol. 111, pp. 767–776.
VIII. John V. Wehausen, *Transformations in linear topological spaces*, Duke Mathematical Journal, vol. 4, pp. 157–169.

Annals of Mathematics
Vol. 44, No. 1, January, 1943

CONTRIBUTIONS TO THE ANALYTIC THEORY OF CONTINUED FRACTIONS AND INFINITE MATRICES

By E. D. Hellinger and H. S. Wall

(Received June 16, 1942)

1. Introduction

In this paper we study continued fractions of the form

$$\frac{1}{b_1 + z} - \frac{a_1^2}{b_2 + z} - \frac{a_2^2}{b_3 + z} - \cdots, \tag{1.1}$$

in which the a_p are *real and positive*, and the b_p are *complex numbers with nonnegative imaginary parts.* In addition, we are concerned with certain infinite matrices connected with this continued fraction.

In the earlier investigations of continued fractions of this form, beginning with the classical work of Stieltjes [15][1], the b_p have been supposed *real.* For this case, and with some additional restrictions, Stieltjes was able to connect the continued fraction with one or more integrals of the form

$$\int_0^\infty \frac{d\phi(u)}{z + u}. \tag{1.2}$$

For particular cases where these restrictions are relaxed, Van Vleck [16] obtained again a connection with an analogous integral the range of which he had to extend over the whole real axis. Hilbert's [11] famous theory of bounded quadratic forms in which the ideas of Stieltjes are in the background allows immediate application to continued fractions of the form (1.1) and their connection with integrals of the form (1.2) but with a finite range of integration.[2] Grommer [4] showed that the process of Hilbert can be applied to more general cases where the integral extends from $-\infty$ to $+\infty$. A general theory of the continued fraction (1.1) in which the a_p and b_p are real and $a_p \neq 0$ was first developed by Hamburger [5] following the pattern laid down by Stieltjes. At about the same time the general case was treated by several other mathematicians: Hellinger [6] employed Hilbert's theory of infinite linear systems, R. Nevanlinna [12] used methods of function theory and asymptotic series, and Carleman [2] used his theory of integral equations.

In contrast with earlier extensions of Stieltjes' original theory, the problem of the present paper is unsymmetrical from two points of view: the functions which are obtained exist in general only in the upper half-plane, $\mathfrak{J}(z) > 0$, and the infinite matrix J of the system of equations

$$-a_p x_p + (b_{p+1} + z)x_{p+1} - a_{p+1}x_{p+2} = 0,\ p = 0, 1, 2, \cdots,\ (a_0 = 0), \tag{1.3}$$

Presented to the American Mathematical Society, April, 18, 1942.

[1] Numbers in square brackets refer to the bibliography.

[2] Cf. Hellinger-Toeplitz [8].

is not Hermitian. In spite of this asymmetry, we are able to develop a theory analogous to the classical theory. The principal points may be summarized as follows.

A. The nest of circles.[3] Regarding the continued fraction as an infinite product of linear transformations, we find for $\mathfrak{J}(z) > 0$ a nest of circles $K_1(z)$, $K_2(z)$, $K_3(z)$, $\cdots$ each inside the preceding and lying in the lower half-plane, such that the n-th approximant of (1.1) is on $K_n(z)$. We find for the radius $r_n(z)$ of $K_n(z)$ the formula

$$r_n(z) = \frac{1}{2\sum_{p=1}^{n} \mathfrak{J}(b_p + z)\,|\,B_{p-1}(z)\,|^2},$$

(cf. (2.16) and (2.13)) where $x_p = B_{p-1}(z)$ is that solution of (1.3) for which $x_1 = 1$. We have to distinguish two cases according as $r(z) = \lim_{n=\infty} r_n(z)$ is positive (*limit-circle case*) or zero (*limit-point case*). In the latter case, (1.1) converges, while in the former case it may converge or diverge (§2).

B. Theorem of invariability. We show that the distinction between the two cases is independent of the particular value of z. This is accomplished in two different ways. 1. Using methods of function theory, particularly the *Stieltjes-Vitali* theorem, we show that it is sufficient to consider the continued fraction (1.1) for $z = 0$. *In this way we find that in addition to the distinction between the two cases, also the fact of convergence or divergence of* (1.1) *is independent of the particular value of* z (§ 4, 5). 2. Using for the linear difference equations an idea which Weyl [19][4] has applied in similar problems on differential and difference equations, we prove a more general formulation of the theorem of invariability (Theorem 7.1).

C. Reciprocals of the matrix J.[5] The matrix J always has at least one bounded reciprocal. However, while for real b_p the limit-point case is characterized by the fact that there is just one bounded reciprocal, there may be infinitely many for complex b_p. Therefore, we introduce a more restrictive boundedness condition (E-boundedness)[6] and obtain an analogous characterization, for the limit-point case (§6).

D. Asymptotic representation. The nest of circles yields a necessary and sufficient condition for an arbitrary function $f(z)$ to be asymptotically equal (Definition 8.1) to the continued fraction (§8).

---

[3] These circles have been applied for the case of real b_p by Hellinger [6]. Earlier applications in related problems were made by Weyl [18] and Hamburger [5]. Cf. also R. Nevanlinna [12], H. Weyl [19], and Paydon-Wall [13].

[4] The use of this idea which extends our original theorem and simplifies the proof was kindly suggested to us by the referee. Cf. further footnote 15.

[5] This theory has been given for the case of real a_p, b_p by Hellinger [6]; cf. also Beth [1].

[6] This amounts only to introducing weight factors in the sum of squares of the variables in the usual definition of boundedness. This has often been done ever since the start of the theory (cf. e.g. [9], §12). For problems somewhat related to our problem one will find an analogous change of the condition of convergence (boundedness) in Weyl [19], in the space of infinitely many variables as well as in the space of functions.

E. Stieltjes integral representation. If for $\mathfrak{J}(z) > 0$ the values of an analytic function $f(z)$ lie in the circle $K_n(z)$ for every n, then $f(z)$ has an integral representation of the form (1.2) with range of integration in general extended over the entire real axis, $\phi(u)$ being still a real nondecreasing function. Unlike the case where the b_p are real, the moments

$$\int_{-\infty}^{+\infty} u^p d\phi(u), \qquad p = 0, 1, 2, \cdots,$$

do not necessarily exist in the general case (§8).

Part I

Continued Fractions and Linear Transformations in One Variable

2. The nest of circles

A continued fraction

$$\frac{c_1}{d_1} - \frac{c_2}{d_2} - \frac{c_3}{d_3} - \cdots \tag{2.1}$$

may be regarded as an "infinite product" of the linear fractional transformations

$$t_n(w) = c_n/(d_n - w), \qquad n = 1, 2, 3, \cdots.$$

The case where these transformations carry some circular region H into all or a part of itself

$$t_n(H) \subset H \tag{2.2}$$

is of particular interest. For if in this case we put $t_1t_2 \cdots t_n(H) = H_n$, $n = 1, 2, 3, \cdots$, then $H \supset H_1 \supset H_2 \supset H_3 \supset \cdots$, and we are therefore in possession of important facts concerning the values of the "generalized approximants" $t_1t_2 \cdots t_n(w)$ of (2.1) when the parameter w lies in H.

We shall be concerned with the sequence of transformations

$$t_n(w) = \frac{a_{n-1}a_n^{-1}}{a_n^{-1}b_n - w}, \qquad n = 1, 2, 3, \cdots, \tag{2.3}$$

where $a_0 = 1, a_1, a_2, \cdots$ are *positive real numbers*, and $b_1, b_2, b_3, \cdots$ are complex numbers

$$b_n = \mathfrak{R}(b_n) + i\mathfrak{J}(b_n) \tag{2.4}$$

with nonnegative imaginary part

$$\mathfrak{J}(b_n) = \beta_n \geqq 0, \qquad n = 1, 2, 3, \cdots. \tag{2.5}$$

If H is the lower half-plane, $\mathfrak{J}(w) \leqq 0$, then under these conditions the transformation $t = t_n(w)$ from (2.3) satisfies the condition (2.2) because $a_{n-1}a_n^{-1}$ is real and positive and $\mathfrak{J}(a_n^{-1}b_n) \geqq 0$.

The continued fraction generated by these transformations is

$$(2.6) \qquad \frac{a_1^{-1}}{a_1^{-1}b_1} - \frac{a_1 a_2^{-1}}{a_2^{-1}b_2} - \frac{a_2 a_3^{-1}}{a_3^{-1}b_3} - \cdots .$$

The notation has been chosen in such a way as to simplify and lend symmetry to the formulas which will be employed. The continued fraction is of course equivalent to

$$\frac{1}{b_1} - \frac{a_1^2}{b_2} - \frac{a_2^2}{b_3} - \cdots .$$

We may write $t_1 t_2 \cdots t_n(w)$ in the form

$$(2.7) \qquad t = t_1 t_2 \cdots t_n(w) = \frac{A_n - A_{n-1}w}{B_n - B_{n-1}w}, \qquad n = 1, 2, 3, \cdots .$$

The A_n, B_n are independent of w and satisfy the recurrence formulas

$$(2.8) \qquad \begin{aligned} a_{n+1}A_{n+1} &= b_{n+1}A_n - a_n A_{n-1}, \\ a_{n+1}B_{n+1} &= b_{n+1}B_n - a_n B_{n-1}, \end{aligned} \qquad n = 0, 1, 2, \cdots ,$$

where

$$(2.9) \qquad A_{-1} = -1, \quad B_{-1} = 0, \quad A_0 = 0, \quad B_0 = 1, \quad A_1 = a_1^{-1}, \quad B_1 = a_1^{-1}b_1.$$

The determinant of the transformation (2.7) may be readily shown to be given by

$$(2.10) \qquad A_n B_{n-1} - A_{n-1}B_n = a_n^{-1}, \qquad n = 1, 2, 3, \cdots .$$

The inverse transformation is

$$(2.11) \qquad w = t_n^{-1} t_{n-1}^{-1} \cdots t_1^{-1}(t) = \frac{A_n - B_n t}{A_{n-1} - B_{n-1}t}.$$

The quotient A_n/B_n is the n-th approximant of the continued fraction (2.6). By (2.7) we have: $A_n/B_n = t_1 t_2 \cdots t_n(0) = t_1 t_2 \cdots t_{n+1}(\infty)$.

From the second recursion formula (2.8) we get

$$a_n B_n \bar{B}_{n-1} = b_n \mid B_{n-1} \mid^2 - a_{n-1}B_{n-2}\bar{B}_{n-1}, \qquad n = 1, 2, 3, \cdots ,$$

or if we put

$$(2.12) \qquad q_n = \Im(B_n \bar{B}_{n-1}), \qquad n = 1, 2, 3, \cdots ,$$

and consider only the imaginary part, we find

$$a_n q_n = \beta_n \mid B_{n-1} \mid^2 + a_{n-1}q_{n-1}, \qquad n = 1, 2, 3, \cdots ,$$

and therefore, using (2.9),

$$(2.13) \qquad a_n q_n = \sum_{p=1}^{n} \beta_p \mid B_{p-1} \mid^2, \qquad n = 1, 2, 3, \cdots .$$

If we recall that $B_0 = 1$ and that $\beta_p \geqq 0$ and if we assume in addition that

$$\mathfrak{J}(b_1) = \beta_1 > 0, \tag{2.14}$$

we can conclude that $q_n > 0$. Therefore, (2.12) shows that $B_n \neq 0$.

Next we determine the region H_n into which the lower half-plane $\mathfrak{J}(w) \leqq 0$ is carried by one of the transformations (2.7). Since $B_n \neq 0$, $B_{n-1} \neq 0$, this transformation maps the real w-axis into a proper circle K_n of the t-plane which bounds H_n. Inasmuch as it maps the point $w = B_n/B_{n-1}$ into $t = \infty$ and the center $t = C_n$ of K_n can be produced by inversion of $t = \infty$ in K_n, then C_n must correspond under this transformation to the reflection $\bar{w} = \bar{B}_n/\bar{B}_{n-1}$ of $w = B_n/B_{n-1}$ in the real w-axis:

$$C_n = \frac{A_n \bar{B}_{n-1} - A_{n-1}\bar{B}_n}{B_n \bar{B}_{n-1} - B_{n-1}\bar{B}_n}. \tag{2.15}$$

Since $t_1 t_2 \cdots t_n(0) = A_n/B_n$ lies upon K_n we then find that the radius of K_n is

$$r_n = \left| \frac{A_n}{B_n} - C_n \right| = \frac{1}{2a_n q_n}. \tag{2.16}$$

In particular, the circle K_1 has the center $C_1 = -i/2\beta_1$ and the radius $r_1 = 1/2q_1 = 1/2\beta_1$, and therefore is tangent to the real axis from below at the origin. Since we have proved $H_1 \supset H_2 \supset H_3 \supset \cdots$ we conclude that for all points of these domains

$$-\frac{1}{\beta_1} \leqq \mathfrak{J}(w) \leqq 0, \qquad |w| \leqq \frac{1}{\beta_1}. \tag{2.17}$$

We have to distinguish two cases:

Case I. The limit-point case. The circular regions H_n have one and only one point f in common; the radius r_n of the circle K_n has the limit 0 for $n = \infty$; the infinite series

$$\sum_{p=1}^{\infty} \beta_p \, | B_{p-1} |^2 \tag{2.18}$$

diverges.

Case II. The limit-circle case. The circular regions H_n have a circular region in common; the circle bounding this common circular region has radius

$$r = \lim_{n=\infty} r_n > 0;$$

the infinite series (2.18) converges.

In Case I we have, *uniformly* for all w in H:

$$\lim_{n=\infty} t_1 t_2 \cdots t_n(w) = f.$$

Inasmuch as $w = 0$ is in H it follows that $\lim_{n=\infty}(A_n/B_n) = f$. Therefore, since the denominators B_n are different from 0, the continued fraction (2.6) is convergent in this case, and its value is f. The continued fraction will be called

completely convergent[7] in Case I. In Case II the continued fraction may converge or diverge; if it converges we shall say that it is *simply convergent.*

3. Complete convergence

We shall give two theorems which furnish sufficient conditions for the complete convergence of the continued fraction (2.6).

THEOREM 3.1. *If* lim inf a_n *is finite, and* $\mathfrak{J}(b_n) \geqq k > 0$, $n = 1, 2, 3, \cdots$, *where* k *is a positive constant, then the continued fraction* (2.6) *is completely convergent.*[8]

PROOF. We prove the equivalent statement: If the series (2.18) converges and if

$$\mathfrak{J}(b_p) = \beta_p \geqq k > 0, \tag{3.1}$$

then $\lim a_n = \infty$. Indeed, from the convergence of (2.18) it follows that $\lim_{p=\infty}\beta_p \mid B_{p-1} \mid^2 = 0$, so that, on account of (3.1), $\lim_{p=\infty}B_{p-1} = 0$, and

$$\lim_{p=\infty} \mathfrak{J}(B_p \bar{B}_{p-1}) = 0. \tag{3.2}$$

On the other hand, we conclude from (2.13) and (3.1) that

$$a_n \mathfrak{J}(B_n \bar{B}_{n-1}) = \sum_{p=1}^{n} \beta_p \mid B_{p-1} \mid^2 \geqq k \mid B_0 \mid^2 = k > 0,$$

which, with (3.2), gives $\lim_{n=\infty}a_n = \infty$.

The second theorem concerns the continued fraction

$$\frac{1}{b_1} - \frac{1}{b_2} - \frac{1}{b_3} - \cdots \tag{3.3}$$

which is obtained by letting $a_n = 1$, $n = 1, 2, 3, \cdots$, in (2.6). As a matter of fact, this is no essential specialization. We note first of all that (3.3) necessarily diverges if all the b_n with odd subscripts vanish. For in this case $B_1 = B_3 = B_5 = \cdots = 0$. Suppose that $b_1 = b_3 = b_5 = \cdots = b_{2n-1} = 0$, $b_{2n+1} \neq 0$. Then, inasmuch as

$$t_1 t_2 \cdots t_{2n+1}(w) = -b_2 - b_4 - \cdots - b_{2n} + \frac{1}{b_{2n+1} - w},$$

it is clear that (3.3) is equivalent to a continued fraction of the same form but with the term $(-b_2 - b_4 - \cdots - b_{2n})$ added on, and with the first partial denominator different from 0.

THEOREM 3.2. *Suppose that* $b_1 \neq 0$, $\mid \mathfrak{R}(b_p) \mid \leqq k\mathfrak{J}(b_p)$, $p = 1, 2, 3, \cdots$, *where* k *is a positive constant, so that* $\mathfrak{J}(b_1) = \beta_1 > 0$. *Then the continued fraction*

[7] For real w this notion has been introduced by Hamburger [5].

[8] See Hellinger [6] for the case where $\mathfrak{J}(b_n)$ is independent of n.

(3.3) *is completely convergent if the series* $\sum |b_p|$ *diverges, and is divergent if this series converges.*[9]

PROOF. If $\sum |b_p|$ diverges, then the inequality $|b_p| \leqq (1 + k)\beta_p$, $\beta_p = \mathfrak{J}(b_p)$, shows that $\sum \beta_p$ diverges. Since $B_p \neq 0$, $p = 0, 1, 2, \cdots$, we conclude from (2.8) that

$$\left|\frac{B_p}{B_{p-2}}\right| = \left|\frac{b_p B_{p-1} - B_{p-2}}{B_{p-2}}\right| \leqq 1 + \left|\frac{b_p B_{p-1}}{B_{p-2}}\right|$$

$$= 1 + \frac{|b_p| \cdot |B_{p-1}|^2}{|B_{p-1}\bar{B}_{p-2}|} \leqq 1 + \frac{(1 + k)\beta_p |B_{p-1}|^2}{q_{p-1}}.$$

Therefore, since by (2.13) $\{q_n\}$ is a monotone nondecreasing sequence and $q_1 = \beta_1 > 0$:

$$\left|\frac{B_p}{B_{p-2}}\right| \leqq 1 + \frac{(1 + k)}{\beta_1}(q_p - q_{p-1}) \leqq 1 + \frac{(1 + k)}{\beta_1}(q_p - q_{p-2}),$$

$$p = 2, 3, 4, \cdots,$$

so that

$$|B_p| \leqq |B_{p-2}| e^{c_1} (q_p - q_{p-2}) \leqq c_2 e^{c_1 q_p},$$

where c_1, c_2 are positive constants. Now $|B_p B_{p-1}| \geqq q_p > \beta_1$, so that

$$|B_{p-1}| \geqq \frac{\beta_1}{|B_p|} \geqq c_3 e^{-c_1 q_p},$$

c_3 being a positive constant, or $\beta_p |B_{p-1}|^2 e^{2c_1 q_p} \geqq c_3^2 \beta_p$. Summing over p from 1 to n we get

$$\sum_{p=1}^{n} \beta_p |B_{p-1}|^2 e^{2c_1 q_p} \geqq c_3^2 \sum_{p=1}^{n} \beta_p ;$$

or, since $q_p \geqq q_{p-1}$:

$$q_n e^{2c_1 q_n} = e^{2c_1 q_n} \sum_{p=1}^{n} \beta_p |B_{p-1}|^2 \geqq c_3^2 \sum_{p=1}^{n} \beta_p .$$

Inasmuch as $\sum \beta_p$ diverges, we conclude that $\lim_{n=\infty} q_n = \infty$, and therefore Case I holds.

If $\sum |b_p|$ converges the continued fraction diverges by a well-known theorem (cf., e.g. Perron [14] p. 235).

4. Theorem of invariability

We consider now the continued fraction obtained from (2.6) by replacing b_p by $b_p + z$, $p = 1, 2, 3, \cdots$, namely:

$$\text{(4.1)} \qquad \frac{a_1^{-1}}{a_1^{-1}(b_1 + z)} - \frac{a_1 a_2^{-1}}{a_2^{-1}(b_2 + z)} - \frac{a_2 a_3^{-1}}{a_3^{-1}(b_3 + z)} - \cdots .$$

[9] This theorem, except for the notion of complete convergence, has been given first by Van Vleck [17], p. 229, Theorem 6.

We keep the conditions

$$a_n > 0,\ \mathfrak{I}(b_n) = \beta_n \geqq 0, \qquad n = 1, 2, 3, \cdots, \tag{4.2}$$

and drop the condition $\mathfrak{I}(b_1) = \beta_1 > 0$. If we suppose that

$$\mathfrak{I}(z) = y > 0 \tag{4.3}$$

then the preceding theory will hold for (4.1) because $\mathfrak{I}(b_1 + z) = \beta_1 + y \geqq y > 0$. We shall use the notation of §2 except that dependence upon z will be indicated in the customary manner. In particular, the denominators will be denoted by $B_n(z)$; they satisfy the recursion formulas (2.8) which now read

$$-a_nB_{n-1}(z) + (b_{n+1} + z)B_n(z) - a_{n+1}B_{n+1}(z) = 0, \qquad n = 0, 1, 2, \cdots. \tag{4.4}$$

This is a system of infinitely many linear equations in $B_0(z)$, $B_1(z)$, $B_2(z)$, $\cdots$ with the following matrix of coefficients:

$$J = \begin{pmatrix} b_1 + z, & -a_1, & 0, & 0, & \cdots \\ -a_1, & b_2 + z, & -a_2, & 0, & \cdots \\ 0, & -a_2, & b_3 + z, & -a_3, & \cdots \\ 0, & 0, & -a_3, & b_4 + z, & \cdots \\ & \cdots & \cdots & \cdots & \end{pmatrix}. \tag{4.5}$$

Matrices of this form in which the coefficients $a_n \neq 0$ have been considered in the theory of infinite matrices under the name of *J-matrices* [8]. Since the theory of these matrices is equivalent to the theory of the continued fractions (4.1) it is appropriate to call these continued fractions *J-fractions*.

The J-fraction (4.1) is completely convergent for $y > 0$ if and only if the series

$$\sum_{p=1}^{\infty} (\beta_p + y) \mid B_{p-1}(z) \mid^2, \quad \text{where } z = x + iy \text{ and } \beta_p = \mathfrak{I}(b_p), \tag{4.6}$$

diverges. When (4.6) diverges, then the radius $r_n(z)$ of the circle $K_n(z)$ tends to 0 as n tends to ∞. The J-fraction is then convergent for this value of z, and its value is $f(z)$, the point common to all the circles $K_n(z)$. We shall prove that if the series (4.6) diverges for one value of $z = x + iy$, where $y > 0$, then it diverges for all such values of z. This will be accomplished by obtaining a condition for complete convergence which is independent of z, namely:

THEOREM 4.1. *The J-fraction* (4.1) *is completely convergent for* $\mathfrak{I}(z) > 0$ *if and only if at least one of the two infinite series*

$$\sum_{p=1}^{\infty} (1 + \beta_p) \mid B_{p-1}(0) \mid^2, \tag{4.7}$$

or

$$\sum_{p-1}^{\infty} (1 + \beta_p) \mid A_{p-1}(0) \mid^2 \tag{4.8}$$

is divergent.[10]

[10] See Hamburger [5] for the case b_p real.

It is easily seen that the series (4.6) and the series

$$\sum_{p=1}^{\infty} (\beta_p + y) \mid A_{p-1}(z) \mid^2 \tag{4.9}$$

converge or diverge together for $y > 0$. In fact, the ratio of corresponding terms is $\mid A_p(z)/B_p(z) \mid^2$, and there is the inequality

$$\frac{1}{\mid b_1 + z \mid + (a_1^2/y)} \leqq \left| \frac{A_p(z)}{B_p(z)} \right| \leqq \frac{1}{y}, \qquad y > 0. \tag{4.10}$$

The second part of the inequality is contained in (2.17). The first part follows then from the remark that $a_1^{-2}(b_1 + z - [B_n(z)/A_n(z)])$ is an approximant of another J-fraction obtained from (4.1) by advancing all indices by unity.

The determinant formula (2.10) shows that the polynomials $A_n(z)$, $B_n(z)$ do not vanish simultaneously. Therefore we conclude from (4.10) that neither vanishes in the upper half-plane $y > 0$. Now if z_1 is a zero of one of these polynomials and if $y \geqq 0$, the length of the vector $iy - z_1$ must increase as y increases. The same is true of any product of lengths of such vectors. Hence it follows that $\mid A_n(0) \mid < \mid A_n(iy) \mid$, $\mid B_n(0) \mid < \mid B_n(iy) \mid$ if $y > 0$. From these considerations we conclude that if at least one of the series (4.7), (4.8) diverges then the series $\sum_{p=1}^{\infty}(\beta_p + y) \mid B_{p-1}(iy) \mid^2$, i.e. the series (4.6) for $z = iy$, diverges for $y > 0$.

It remains to be shown that the series (4.6) diverges for any $z_0 = x_0 + iy_0$ with $y_0 > 0$. To do this, we show that any two points L_1, L_2 which are inside of every circle $K_n(z_0)$, $n = 1, 2, 3, \cdots$, are identical. In fact, we may select two sequences $\{u_n\}$ and $\{v_n\}$ lying in the lower half-plane such that

$$\lim_{n=\infty} t_1 t_2 \cdots t_n(z_0 ; u_n) = \lim_{n=\infty} \frac{A_n(z_0) - u_n A_{n-1}(z_0)}{B_n(z_0) - u_n B_{n-1}(z_0)} = L_1,$$

$$\lim_{n=\infty} t_1 t_2 \cdots t_n(z_0 ; v_n) = \lim_{n=\infty} \frac{A_n(z_0) - v_n A_{n-1}(z_0)}{B_n(z_0) - v_n B_{n-1}(z_0)} = L_2 .$$

Let G be any bounded closed connected region in the upper half-plane $y > 0$ which contains on its interior the point z_0 and a portion of the positive half of the imaginary axis. The two sequences of rational functions of z,

$$\{t_1 t_2 \cdots t_n(z; u_n)\}, \qquad \{t_1 t_2 \cdots t_n(z; v_n)\} \tag{4.11}$$

are uniformly bounded over G. Hence we may select two subsequences, one from each, which are uniformly convergent over G to analytic limit-functions $f_1(z)$ and $f_2(z)$, respectively. Inasmuch as, for the pure imaginary points of G, $\lim r_n(iy) = 0$ so that $f_1(iy) = f_2(iy)$, it follows that $f_1(z) \equiv f_2(z)$ for all z in G. Therefore $L_1 = L_2$ and consequently (4.6) diverges if $\mathfrak{J}(z) = y > 0$.

We now suppose that both the series (4.7) and (4.8) converge, and shall prove that (4.6) converges for *all* values of z. The plan of the proof is as follows. We shall assign to u_n and v_n in (4.11) particular values in the lower half-plane in such a way that the sequences (4.11) will converge to limit-functions $f_1(z)$ and

$f_2(z)$ where $f_1(z) - f_2(z)$ is nowhere zero. This will of course imply that (4.6) converges for $\Im(z) > 0$. On account of the fact that none of the zeros of $B_p(z)$ are in the upper half-plane, we conclude by similar considerations as used before that (4.6) converges for all z.

On referring to (2.13) we recall that if $y \geqq 0$, then $\Im(B_n\bar{B}_{n-1}) \geqq 0$. Consequently, $\Im(\bar{B}_nB_{n-1}) \leqq 0$. Similarly, $\Im(\bar{A}_nA_{n-1}) \leqq 0$. Hence two sequences of constants lying in the lower half-plane are:

$$u_n = \frac{\bar{A}_n(0)}{\bar{A}_{n-1}(0)}, \qquad v_n = \frac{\bar{B}_n(0)}{\bar{B}_{n-1}(0)}, \qquad n = 1, 2, 3, \cdots.$$

These we shall use in forming the sequences (4.11). Put

$$(4.12)\quad \begin{aligned} &\text{(i)} & P_n(z) &= a_n(A_n(z)\bar{A}_{n-1}(0) - A_{n-1}(z)\bar{A}_n(0)),\\ &\text{(ii)} & Q_n(z) &= a_n(B_n(z)\bar{A}_{n-1}(0) - B_{n-1}(z)\bar{A}_n(0)),\\ &\text{(iii)} & U_n(z) &= a_n(A_n(z)\bar{B}_{n-1}(0) - A_{n-1}(z)\bar{B}_n(0)),\\ &\text{(iv)} & V_n(z) &= a_n(B_n(z)\bar{B}_{n-1}(0) - B_{n-1}(z)\bar{B}_n(0)),\end{aligned}$$

and (4.11) are then the sequences $\{P_n(z)/Q_n(z)\}$, $\{U_n(z)/V_n(z)\}$, respectively. Inasmuch as the numbers u_n and v_n lie in the lower half-plane, $P_n(z)/Q_n(z)$ and $U_n(z)/V_n(z)$ must lie in the circle $K_n(z)$. In view of the preceding remarks, the proof of Theorem 4.1* will be complete when we have proved the following theorem.

THEOREM 4.2. *If the series* (4.7) *and* (4.8) *both converge, then there exist four entire functions* $p(z)$, $q(z)$, $u(z)$, $v(z)$ *such that*

$$(4.13)\qquad p(z)v(z) - u(z)q(z) = 1,$$

and such that

$$(4.14)\quad \begin{aligned} \lim_{n=\infty} P_n(z) &= p(z), & \lim_{n=\infty} Q_n(z) &= q(z),\\ \lim_{n=\infty} U_n(z) &= u(z), & \lim_{n=\infty} V_n(z) &= v(z),\end{aligned}$$

uniformly over every bounded region of the z-plane.

PROOF.[11] From the determinant formula (2.10) we find by (4.12) that $P_n(z)V_n(z) - U_n(z)Q_n(z) = 1$, so that (4.13) will follow from (4.14). In the proof of (4.14) we shall use for the sake of brevity the notations:

$$\begin{aligned}\lambda_n(z) &= (z + 2i\beta_{n+1})\bar{A}_n(0)^2\\ \mu_n(z) &= (z + 2i\beta_{n+1})\bar{B}_n(0)^2\\ \nu_n(z) &= (z + 2i\beta_{n+1})\bar{A}_n(0)\bar{B}_n(0).\end{aligned}$$

[11] The essential idea of the proof is contained in several earlier investigations on continued fractions; cf., for instance, O. Perron [14], p. 235.

Since the series (4.7) and (4.8) converge and since $\beta_{n+1} \geqq 0$ it follows that the series

$$\sum \lambda_p(z), \qquad \sum \mu_p(z), \qquad \sum \nu_p(z),$$

and also the infinite product

$$\prod (1 - \nu_p(z))$$

converge absolutely and uniformly over any bounded closed region G of the z-plane. Moreover, there exists an N depending only upon G such that

$$\pi_n(z) = \prod_{p=N}^{N+n} (1 - \nu_p(z)) \neq 0, \qquad n = 0, 1, 2, \cdots, \quad z \text{ in } G.$$

On eliminating $A_{n+1}(z)$, $A_n(z)$, $A_{n-1}(z)$ from the recurrence formula (cf. (2.8))

$$a_{n+1}A_{n+1}(z) = (b_{n+1} + z)A_n(z) - a_nA_{n-1}(z),$$

and the equations (i) (written for n and $n + 1$) and (iii) of (4.12), we get a linear relation among $P_{n+1}(z)$, $U_n(z)$ and $P_n(z)$. Analogously we get all the following four identities

$$\text{(4.15)} \quad \begin{array}{ll} \text{(i)} & P_{n+1}(z) = \lambda_n(z)U_n(z) + (1 - \nu_n(z))P_n(z), \\ \text{(ii)} & Q_{n+1}(z) = \lambda_n(z)V_n(z) + (1 - \nu_n(z))Q_n(z), \\ \text{(iii)} & (1 - \nu_n(z))U_{n+1}(z) = -\mu_n(z)P_{n+1}(z) + U_n(z), \\ \text{(iv)} & (1 - \nu_n(z))V_{n+1}(z) = -\mu_n(z)Q_{n+1}(z) + V_n(z). \end{array}$$

If in (i) and (iii) we replace n by $N + n$ where N is the index introduced before, and if we use the notation

$$P_n^*(z) = \frac{P_{N+n}(z)}{\pi_{n-1}(z)}, \qquad U_n^*(z) = \pi_{n-1}(z)U_{N+n}(z),$$

$$b_n^*(z) = \frac{\lambda_{N+n}(z)}{\pi_{n-1}(z)\pi_n(z)}, \qquad c_n^*(z) = -\mu_{N+n}(z)\pi_{n-1}(z)\pi_n(z),$$

these relations become

$$\text{(4.16)} \quad \begin{array}{l} P_{n+1}^*(z) = b_n^*(z)U_n^*(z) + P_n^*(z), \\ U_{n+1}^*(z) = c_n^*(z)P_{n+1}^*(z) + U_n^*(z), \end{array} \qquad n = 1, 2, 3, \cdots.$$

By the remark at the beginning, the series $\sum b^*(z)$ and $\sum c^*(z)$ converge absolutely and uniformly over G, and hence there exists a finite number M_1 such that, for all z in G:

$$\prod_{p=1}^{\infty} (1 + | b_p^*(z) |) < M_1, \qquad \prod_{p=1}^{\infty} (1 + | c_p^*(z) |) < M_1.$$

Now, if $| U_1^*(z) | \leqq M_2$, $| P_1^*(z) | \leqq M_2$ over G, we have by (4.16):

$$| P_2^*(z) | \leqq | b_1^*(z) | \cdot | U_1^*(z) | + | P_1^*(z) | \leqq (1 + | b_1^*(z) |)M_2,$$

$$| U_2^*(z) | \leqq | c_1^*(z) | \cdot | P_2^*(z) | + | U_1^*(z) | \leqq (1 + | b_1^*(z) |)(1 + | c_1^*(z) |)M_2.$$

Continuing this procedure we find

$$| P_n^*(z) | \leqq M, \qquad | U_n^*(z) | \leqq M, \qquad n = 1, 2, 3, \cdots, z \text{ in } G,$$

where $M = M_2 M_1^2$.

Furthermore, by (4.16) we have $P_{n+1}^*(z) = P_1^*(z) + \sum_{p=1}^n b_p^*(z) U_p^*(z)$ and therefore

$$\lim_{n=\infty} P_{N+n+1}(z) = \lim_{n=\infty} \pi_n(z) P_{n+1}^*(z) = [\lim_{n=\infty} \pi_n(z)] \left[P_1^*(z) + \sum_{p=1}^{\infty} b_p^*(z) U_p^*(z) \right],$$

uniformly over G. This establishes the first limit in (4.14); the proof of the other limits can be made in the same way.

We have thus established in Theorem 4.1 a condition for the complete convergence of the J-fraction which does not depend upon z. This means, in fact, that when at least one of the constant term series (4.7) and (4.8) diverges, then the J-fraction (4.1) converges if $\mathfrak{J}(z) > 0$. Since the sequence of approximants is uniformly bounded for $\mathfrak{J}(z) \geqq k > 0$, the convergence is uniform over any finite region in this domain. We therefore have the following result.

THEOREM 4.3. *If at least one of the series* (4.7), (4.8) *diverges, the J-fraction* (4.1) *converges and represents an analytic function of z for* $\mathfrak{J}(z) > 0$.

We have also obtained in Theorem 4.2 the means which will enable us in the next section to answer completely the questions of convergence of the J-fraction and nature of the limit-function when both the series (4.7), (4.8) converge.

5. Simple convergence

On eliminating $A_{n-1}(z)$ from (i) and (iii) of (4.12) and $B_{n-1}(z)$ from (ii) and (iv), we now obtain the formula

$$\frac{A_n(z)}{B_n(z)} = \frac{P_n(z) - s_n U_n(z)}{Q_n(z) - s_n V_n(z)},$$

where $s_n = \bar{A}_n(0)/\bar{B}_n(0)$. When the series (4.7) and (4.8) both converge and $\lim_{n=\infty} s_n = s$ is finite, then the numerator and denominator converge to $p(z) - su(z)$ and $q(z) - sv(z)$, respectively. Since by (4.13) these two functions cannot vanish simultaneously, and $q(z) - sv(z)$ is not identically 0, being different from 0 for $\mathfrak{J}(z) > 0$, we conclude that the J-fraction converges to the quotient $[p(z) - su(z)]/[q(z) - sv(z)]$, which is a meromorphic function of z. The convergence is uniform in any closed bounded region containing no poles of the limit function. Similarly, if $\lim_{n=\infty} s_n = \infty$, we conclude that the J-fraction converges in like manner to the meromorphic function $u(z)/v(z)$. If the sequence $\{s_n\}$ has more than one limit-point (one of which may be ∞) it is easily seen with the aid of (4.13) that the J-fraction (4.1) is divergent for all values of z. These statements contain the following theorem.

THEOREM 5.1. *In case both the series* (4.7) *and* (4.8) *converge, then the convergence of the J-fraction* (4.1) *or of its reciprocal for a single value of z implies the convergence of the J-fraction or its reciprocal for any value of z. The value of the*

J-fraction is a meromorphic function of z, namely, in terms of the entire functions of (4.14),

$$\frac{p(z) - su(z)}{q(z) - sv(z)} \quad \text{if} \quad \lim_{n=\infty} \frac{\bar{A}_n(0)}{\bar{B}_n(0)} = s \quad \text{is finite};$$

$$\frac{u(z)}{v(z)} \quad \text{if} \quad \lim_{n=\infty} \frac{\bar{B}_n(0)}{\bar{A}_n(0)} = 0.$$

The J-fraction is not completely convergent in this case.

Part II

Continued Fractions and Linear Transformations in Infinitely Many Variables

6. J-matrices

In contrast with the preceding theory there is another theory in which the continued fraction appears in connection with a single linear transformation in infinitely many variables, namely:

$$\begin{aligned} (b_1 + z)x_1 - \quad a_1x_2 \quad\quad\quad\quad\quad\quad &= y_1, \\ -a_1x_1 + (b_2 + z)x_2 - \quad a_2x_3 \quad\quad\quad &= y_2, \\ -a_2x_2 + (b_3 + z)x_3 - a_3x_4 &= y_3, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{6.1}$$

which carries the point $x = (x_1, x_2, x_3, \cdots)$ into the point $y = (y_1, y_2, y_3, \cdots)$. The matrix of this transformation is the *J-matrix* (4.5). For the present we allow the coefficients to be arbitrary complex numbers, the a_p being of course different from 0. If $y_1 = 1$, $y_n = 0$, $n > 1$, the equations (6.1) may be written as

$$x_1 = \cfrac{1}{b_1 + z - \cfrac{a_1 x_2}{x_1}}, \quad \frac{x_2}{x_1} = \cfrac{a_1}{b_2 + z - \cfrac{a_2 x_3}{x_2}}, \quad \frac{x_3}{x_2} = \cfrac{a_2}{b_3 + z - \cfrac{a_3 x_4}{x_3}}, \quad \cdots, \tag{6.2}$$

and consequently x_1 is formally equal to the J-fraction:

$$\frac{1}{b_1 + z} - \frac{a_1^2}{b_2 + z} - \frac{a_2^2}{b_3 + z} - \cdots. \tag{6.3}$$

If the y_p are arbitrary, one could try to express a solution of (6.1) in the form

$$x_p = \sum_{q=1}^{\infty} \rho_{pq} y_q, \qquad p = 1, 2, 3, \cdots. \tag{6.4}$$

Here, ρ_{11} would be formally equal to the J-fraction (6.3). A matrix (ρ_{pq}) with this property is called a *right reciprocal* of the J-matrix (4.5). On substituting (6.4) into (6.1) we find the relations

$$-a_{p-1}\rho_{p-1,q} + (b_p + z)\rho_{p,q} - a_p\rho_{p+1,q} = \delta_{p,q}, \qquad p, q = 1, 2, 3, \cdots, \tag{6.5}$$

where a_0 is to be set equal to 0, and $\delta_{p,q}$ is equal to 0 or 1 according as $p \neq q$ or $p = q$. Since $a_p \neq 0$, $p = 1, 2, 3, \cdots$, it follows that, for a fixed q, if $\rho_{1,q}$ is chosen arbitrarily, then $\rho_{p,q}$, $p = 2, 3, 4, \cdots$, are uniquely determined. Therefore, there are infinitely many different right reciprocals.

An essential relationship to the J-fraction can be expected only for those reciprocals which belong to certain restricted classes. The most important class is given by the following definition:[12]

DEFINITION 6.1. *The matrix* (k_{pq}) *and the bilinear form in infinitely many variables*

$$K(\xi, \eta) = \sum_{p,q=1}^{\infty} k_{pq} \xi_p \eta_q$$

is said to be bounded if there exists a fixed number M *such that for all values of the* ξ_p *and* η_q *and for all integers* n

$$\left| \sum_{p,q=1}^{n} k_{pq} \xi_p \eta_q \right| \leqq M \cdot \left(\sum_{p=1}^{n} |\xi_p|^2 \right)^{1/2} \cdot \left(\sum_{q=1}^{n} |\eta_q|^2 \right)^{1/2}. \tag{6.6}$$

The explicit formulas for the $\rho_{p,q}$ in terms of the arbitrary $\rho_{1,q}$ and the polynomials $B_p(z)$ and $A_p(z)$ are:

$$\rho_{p,q} = \begin{cases} \rho_{1,q} B_{p-1}(z), & p = 1, 2, 3, \cdots, q; \\ \rho_{1,q} B_{p-1}(z) + A_{q-1}(z) B_{p-1}(z) - A_{p-1}(z) B_{q-1}(z), & \\ & p = q + 1, q + 2, \cdots. \end{cases} \tag{6.7}$$

This may be readily verified by comparing the recurrence formulas for the $A_p(z)$ and $B_p(z)$ with (6.5).

If one introduces new arbitrary functions $w_q(z)$ by means of the equations

$$\rho_{1,q} = B_{q-1}(z) w_q(z) - A_{q-1}(z), \qquad q = 1, 2, 3, \cdots,$$

then the formulas (6.7) take the form

$$\rho_{p,q}(z) = \begin{cases} B_{p-1}(z) B_{q-1}(z) \left(w_q(z) - \dfrac{A_{q-1}(z)}{B_{q-1}(z)} \right), & p = 1, 2, 3, \cdots, q; \\[2ex] B_{p-1}(z) B_{q-1}(z) \left(w_q(z) - \dfrac{A_{p-1}(z)}{B_{p-1}(z)} \right), & p = q + 1, q + 2, \cdots. \end{cases} \tag{6.8}$$

From (6.8) one sees immediately that the matrix (ρ_{pq}) is symmetric if and only if $w_1 = w_2 = w_3 = \cdots$; i.e. if and only if[13]

$$\rho_{n+1,q}/\rho_{n,q} = v_n, \qquad n \geqq q,$$

where v_n is independent of q, $(q, n = 1, 2, 3, \cdots)$.

[12] For the theory of bounded matrices see, for instance, Hellinger-Toeplitz [9], [10]. Cf. also H. T. Davis [3].

[13] See Hellinger [6]. Following the procedure which Weyl [18] had applied to boundary value problems of ordinary differential equations, Hellinger used a real parameter t and considered for $q = 1$ the equation $\rho_{n+1,q} = t\rho_{n,q}$ as a boundary condition. Beth [1] did the same for $q > 1$.

We shall suppose now, as in Part I, *that the a_p are real and positive and that the b_p satisfy the condition* $\mathfrak{I}(b_p) \geqq 0$, $p = 1, 2, 3, \cdots$. The next theorem gives a necessary and sufficient condition for the general right reciprocal of J to be bounded in the *limit-circle case* (cf. §2).

THEOREM 6.1. *The right reciprocal* (6.7) *is bounded in the limit-circle case if and only if the series*

$$\sum_{q=1}^{\infty} | \rho_{1,q} |^2 \tag{6.9}$$

is convergent.

PROOF. The necessity of the condition follows from the fact that in a bounded matrix the moduli of the elements of any row have a convergent sum of squares. If we recall that in the limit-circle case the series

$$\sum_{p=1}^{\infty} | A_{p-1}(z) |^2, \qquad \sum_{p=1}^{\infty} | B_{p-1}(z) |^2 \tag{6.10}$$

converge for all values of z, then we find from (6.7) that the convergence of (6.9) implies the convergence of the double series $\sum_{p,q=1}^{\infty} | \rho_{p,q}(z) |^2$. Consequently, the matrix $(\rho_{p,q})$ is not only bounded but is even completely continuous. We therefore have

COROLLARY 6.1. *In the limit-circle case, any right reciprocal* (6.7) *which is bounded is also completely continuous.*

It is possible for the series (6.10) to converge also in the limit-point case. Then there will be infinitely many bounded reciprocals. However, we can modify the condition of boundedness so that it is satisfied by only one such reciprocal. Definition 6.1 introduces an upper bound M for the values of the forms in (6.6), for $n = 1, 2, 3, \cdots$, as the variables run over the *spheres* $\sum_{p=1}^{n} | \xi_p |^2 \leqq 1$. Analogously, we can consider an upper bound as the variables run over the *ellipsoids* $\sum_{p=1}^{n} | \xi_p |^2(1 + \beta_p)^{-1} \leqq 1$, and accordingly we make the following definition:

DEFINITION 6.2. *The matrix (ρ_{pq}) will be said to be E-bounded if there exists a fixed number M such that for all values ξ_p, η_q, and for all integers n:*

$$\left| \sum_{p,q=1}^{n} \rho_{p,q} \xi_p \eta_q \right| \leqq M \cdot \left(\sum_{p=1}^{n} \frac{| \xi_p |^2}{1 + \beta_p} \right)^{1/2} \cdot \left(\sum_{q=1}^{n} \frac{| \eta_q |^2}{1 + \beta_q} \right)^{1/2}, \qquad \beta_p \geqq 0. \tag{6.11}$$

This is equivalent to saying that the matrix $((1 + \beta_p)^{1/2}(1 + \beta_q)^{1/2}\rho_{pq})$ is bounded in the sense of Definition 6.1, and hence the theorems about bounded matrices can be extended immediately to E-bounded matrices.

THEOREM 6.2. *In the limit-point case, let $f(z)$ be the analytic function whose value for every z with $\mathfrak{I}(z) > 0$ is common to all the circles $K_n(z)$. Then the formula* (6.8) *with $w_q(z) = f(z)$, $q = 1, 2, 3, \cdots$, gives the unique E-bounded right reciprocal, which is simultaneously the unique left reciprocal.*

Proof. We show first that if $w_q(z) = f(z)$ for all q and for $\mathfrak{I}(z) > 0$, then the matrix (ρ_{pq}) is E-bounded. If we suppose only $w_1(z) = w_2(z) = \cdots = w_n(z)$, where n is a fixed index, then from (6.8) it follows that

$$\frac{\rho_{n+1,q}}{\rho_{n,q}} = \frac{A_n(z) - w_1(z)B_n(z)}{A_{n-1}(z) - w_1(z)B_{n-1}(z)}, \qquad q = 1, 2, 3, \cdots, n. \tag{6.12}$$

Hence, the quotient $\rho_{n+1,q}/\rho_{n,q} = v_n$ is independent of q:

$$\rho_{n+1,q} = v_n \rho_{n,q}, \qquad q = 1, 2, 3, \cdots, n. \tag{6.13}$$

If we identify (6.12) with (2.11), we conclude at once that $\mathfrak{I}(v_n) \leqq 0$ if and only if $w_1(z)$ has its value in the circle $K_n(z)$.

Let $\xi_1, \xi_2, \cdots, \xi_n$ be arbitrary real numbers. On multiplying both members of (6.13) by ξ_q and summing over q from 1 to n we then have:

$$y_{n+1} = v_n y_n, \qquad y_p = \sum_{q=1}^{n} \rho_{pq} \xi_q .$$

Therefore $y_{n+1}\bar{y}_n = v_n \mid y_n \mid^2$ and consequently

$$\mathfrak{I}(y_{n+1}\bar{y}_n) \leqq 0 \tag{6.14}$$

if $w_1(z)$ is in $K_n(z)$.

We now multiply the equation (6.5) by ξ_q and sum over q from 1 to n. This gives the equation

$$L_p(y) \equiv -a_{p-1}y_{p-1} + (b_p + z)y_p - a_p y_{p+1} = \xi_p, \qquad p = 1, 2, 3, \cdots, \quad a_0 = 0. \tag{6.15}$$

Now, one has immediately the identity[14]

$$\sum_{p=1}^{n} (\bar{y}_p L_p(y) - y_1 \overline{L_p(y)}) = 2i \left(\sum_{p=1}^{n} (\beta_p + \mathfrak{I}(z)) | y_p |^2 - a_n \mathfrak{I}(y_{n+1}\bar{y}_n) \right)$$

and by (6.15) we get

$$a_n \mathfrak{I}(y_{n+1}\bar{y}_n) = \sum_{p=1}^{n} (\beta_p + \mathfrak{I}(z)) | y_p |^2 + \sum_{p=1}^{n} \xi_p \mathfrak{I}(y_p). \tag{6.16}$$

Considering now the quadratic form

$$R_n(\xi, \xi) = \sum_{p,q=1}^{n} \rho_{pq} \xi_p \xi_q = \sum_{p=1}^{n} y_p \xi_p ,$$

[14] This identity is analogous to one of the so-called "Green's formulas" in the theory of differential equations. It may be emphasized that if we regard $\rho_{11} = w$ as a complex variable, then (6.16) along with (6.14) gives the inequality defining our nest of circular regions. This is the method which was used by Weyl [18] for differential equations, and by Hellinger [6] and Beth [1] for real J-fractions. The method has been used also for other problems, cf., for instance, Weyl [19].

we have by Schwarz's inequality

$$|R_n(\xi,\xi)|^2 \leqq \sum_{p=1}^{n} \frac{\xi_p^2}{\beta_p + \Im(z)} \cdot \sum_{p=1}^{n} (\beta_p + \Im(z)) |y_p|^2,$$

and, using (6.16) and (6.14):

$$|R_n(\xi,\xi)|^2 \leqq \sum_{p=1}^{n} \frac{\xi_p^2}{\beta_p + \Im(z)} \cdot \left(-\sum_{p=1}^{n} \xi_p \Im(y_p)\right).$$

Inasmuch as $-\sum_{p=1}^{n} \xi_p \Im(y_p) = -\Im(R_n(\xi,\xi)) \leqq |R_n(\xi,\xi)|$, we have: $|R_n(\xi,\xi)| \leqq \sum_{p=1}^{n} \xi_p^2/(\beta_p + \Im(z))$, or

$$\left|\sum_{p,q=1}^{n} \rho_{pq}\xi_p\xi_q\right| \leqq \sum_{p=1}^{n} \frac{\xi_p^2}{\beta_p + \Im(z)} \leqq M_1 \cdot \sum_{p=1}^{n} \frac{\xi_p^2}{1+\beta_p}, \tag{6.17}$$

for a suitably chosen M_1 depending only upon the domain of z. For the related bilinear form we therefore have

$$\begin{aligned} |R_n(\xi,\eta)| &= \tfrac{1}{4}|R_n(\xi+\eta,\xi+\eta) - R_n(\xi-\eta,\xi-\eta)| \\ &\leqq \frac{1}{2}\sum_{p=1}^{n} \frac{\xi_p^2 + \eta_p^2}{\beta_p + \Im(\xi)}. \end{aligned} \tag{6.18}$$

It follows that if $f(z)$ is the value common to all the circles $K_n(z)$, i.e. the value $\lim (A_p(z)/B_p(z))$ of the J-fraction, and if every $w_q(z) = f(z)$, then the matrix (ρ_{pq}) is E-bounded.

It remains to be seen that any other reciprocal given by (6.8) is not E-bounded in the limit-point case. In fact, for any other reciprocal there is at least one z and integers q and $N \geqq q$ such that

$$\left| w_q(z) - \frac{A_{p-1}(z)}{B_{p-1}(z)} \right| \geqq d > 0 \quad \text{for} \quad p \geqq N.$$

Hence we see immediately by the second equation (6.8) that $\sum_p (1+\beta_p) \cdot (1+\beta_q) |\rho_{pq}|^2$ diverges inasmuch as $\sum_p (1+\beta_p) |B_{p-1}(z)|^2$ diverges. This implies that the matrix $((1+\beta_p)^{1/2}(1+\beta_q)^{1/2}\rho_{pq})$ is not bounded, i.e. the matrix (ρ_{pq}) is not E-bounded.

From the inequality (6.18) we shall now derive some estimates for $\rho_{pq} = \rho_{pq}(z)$ which will be useful later on.

THEOREM 6.3. *Let $w_q(z) = f(z)$, $q = 1, 2, 3, \cdots, n$, in (6.8), where $f(z)$ is in the circle $K_n(z)$ for $\Im(z) > 0$. Then*

$$\begin{gathered} |\rho_{pq}(z)| \leqq \frac{1}{(y+\beta_p)^{1/2}(y+\beta_q)^{1/2}} \leqq \frac{1}{y}, \\ z = x + iy, \qquad y > 0, \qquad p, q \leqq n; \end{gathered} \tag{6.19}$$

and

$$\begin{gathered} \rho_{pq}(z) = \frac{\delta_{pq}}{z} + \frac{\theta_{pq}(z)}{zy} G_p, \\ |\theta_{pq}| \leqq 1, \qquad G_p = a_p + |b_p| + a_{p-1}, \qquad p, q < n. \end{gathered} \tag{6.20}$$

PROOF. The inequality (6.19) is an immediate consequence of (6.18) if we specialize the variables such that one $\xi_p^2 = \mathfrak{J}(z) + \beta_p$, one $\eta_q^2 = \mathfrak{J}(z) + \beta_q$, and the other variables are 0. Now by (6.5):

$$\rho_{pq}(z) = \frac{\delta_{pq}}{z} + \frac{a_p\rho_{p+1,q}(z) - b_p\rho_{pq}(z) + a_{p-1}\rho_{p-1,q}(z)}{z}.$$

Applying (6.19) to this identity we obtain (6.20).

7. A general theorem of invariability

We have seen that the behavior of the J-fraction (4.1) is invariant in some respect under change in the particular value of z in the upper half-plane. This has been implied by the fact that if the series (4.6) converges for one z with $\mathfrak{J}(z) > 0$, it converges for all z. We now derive a more general theorem of invariability which covers *entirely arbitrary* J-fractions. The proof reveals in a certain way the inner structure of the theory.

THEOREM 7.1. *Let* (4.1) *be a J-fraction with entirely arbitrary coefficients* $a_p \neq 0$ *and* b_p. *Let* α_p, $p = 1, 2, 3, \cdots$, *be arbitrary real numbers not less than* 1. *Then, if the two series*

$$(7.1)\qquad \sum_{p=1}^{\infty} \alpha_p^2 \,|\, A_{p-1}(z)\,|^2, \qquad \sum_{p=1}^{\infty} \alpha_p^2 \,|\, B_{p-1}(z)\,|^2$$

converge simultaneously for one value of z, they converge for every value of z.

This obviously contains the theorem of invariability of §4. Moreover, it supplements the remark on the limit-point case after Corollary 6.1 in this way: *If the series* (6.10) *both converge for one value of z, then there exist infinitely many bounded reciprocals for every value of z.*

PROOF OF THEOREM 7.1.[15] Write, as in (6.15) but with a_0 now equal to 1,

$$L_p(y) \equiv -a_{p-1}y_{p-1} + (b_p + z)y_p - a_py_{p+1}, \quad p = 1, 2, 3, \cdots, a_0 = 1,$$

and denote by $L_p^*(y)$ the same expression with z replaced by z^*. The solution of the system $L_p(y) = 0$ under the initial conditions $y_0 = -1$, $y_1 = 0$ is $y_p = A_{p-1}(z) = A_{p-1}$, while under the initial conditions $y_0 = 0$, $y_1 = 1$ the solution is $y_p = B_{p-1}(z) = B_{p-1}$. If y_p, y_p^* are arbitrary solutions of the systems $L_p(y) = 0$ and $L_p^*(y) = 0$, respectively, then we obtain immediately the relation:

$$\sum_{p=1}^{n} (y_p^* L_p(y) - y_p L_p^*(y^*)) = y_1 y_0^* - y_0 y_1^*$$
$$- a_n(y_{n+1}y_n^* - y_n y_{n+1}^*) + (z - z^*) \sum_{p=1}^{n} y_p y_p^* = 0.$$

[15] Cf. footnote 4. This proof uses the idea which Weyl [19] has applied in similar problems, namely, to express the relationship between solutions for two different parameter values as a Volterra integral or sum equation. This procedure as well as the procedure used by Weyl [18] and Hellinger [6] may be embraced in a general set-up, if one uses an arbitrary one of the infinitely many reciprocals (6.7) of the J-matrix. Then, the different forms of the proof appear in specializing the reciprocal in different ways, for instance so that it becomes symmetrical ($\rho_{pq} = \rho_{qp}$) or a Volterra form ($\rho_{pq} = 0$ for $p < q$).

In particular, for $y_p = A_{p-1}$ and $y_p = B_{p-1}$ we get

$$y_1^* - a_n(y_n^* A_n - y_{n+1}^* A_{n-1}) + (z - z^*) \sum_{p=1}^{n} y_p^* A_{p-1} = 0,$$

$$y_0^* - a_n(y_n^* B_n - y_{n+1}^* B_{n-1}) + (z - z^*) \sum_{p=1}^{n} y_p^* B_{p-1} = 0,$$

respectively. On multiplying the first of these equations by $-B_{n-1}$, the second by A_{n-1}, and then adding, we get:

$$y_n^* + (z^* - z) \sum_{p=1}^{n-1} (A_{p-1}B_{n-1} - A_{n-1}B_{p-1})y_p^* = y_1^* B_{n-1} - y_0^* A_{n-1}.$$

Therefore, $\zeta_p = \alpha_p y_p^*$ is a solution of the Volterra sum equation:

$$\zeta_p + \sum_{q=1}^{p-1} k_{pq}\zeta_q = g_p, \qquad p = 1, 2, 3, \cdots, \tag{7.2}$$

in which

$$k_{pq} = \alpha_p \alpha_q^{-1}(z^* - z)(A_{q-1}B_{p-1} - A_{p-1}B_{q-1}), \qquad g_p = \alpha_p(y_1^* B_{p-1} - y_0^* A_{p-1}).$$

The proof of the theorem will be complete if we show that $\sum |\zeta_p|^2$ is convergent.

From the convergence of the series (7.1) it follows at once that g_p satisfies the condition:

$$C^2 = \sum_{p=1}^{\infty} |g_p|^2$$

is finite; and that, inasmuch as $\alpha_p \geqq 1$, the double series $\sum |k_{pq}|^2$ converges, so that for r sufficiently large:

$$\epsilon_r = \sum_{q=1}^{\infty} \sum_{p=r}^{\infty} |k_{pq}|^2 < 1. \tag{7.3}$$

We now multiply (7.2) by $\bar{\zeta}_p$ and sum over p from r to m, $m > r$. This gives, if we apply Schwarz's inequality:

$$\begin{aligned} \sum_{p=r}^{m} |\zeta_p|^2 &\leqq C\cdot\left(\sum_{p=r}^{m} |\zeta_p|^2\right)^{1/2} + \sum_{p=r}^{m}\sum_{q=1}^{m} |k_{pq}\zeta_p\zeta_q| \\ &\leqq \left(\sum_{p=r}^{m} |\zeta_p|^2\right)^{1/2}\left(C + \sum_{q=1}^{m} |\zeta_q| \left(\sum_{p=r}^{m} |k_{pq}|^2\right)^{1/2}\right), \end{aligned}$$

and consequently, again using Schwarz's inequality,

$$\left(\sum_{p=r}^{m} |\zeta_p|^2\right)^{1/2} \leqq C + \epsilon_r\left(\sum_{q=1}^{m} |\zeta_q|^2\right)^{1/2} \leqq C + \epsilon_r\left(\sum_{q=1}^{r-1} |\zeta_q|^2\right)^{1/2} + \epsilon_r\left(\sum_{q=r}^{m} |\zeta_q|^2\right)^{1/2},$$

or

$$\left(\sum_{p=r}^{m} |\zeta_p|^2\right)^{1/2}\cdot(1 - \epsilon_r) \leqq C + \epsilon_r\left(\sum_{q=1}^{r-1} |\zeta_q|^2\right)^{1/2}.$$

Hence, by (7.3), the series $\sum |\zeta_p|^2$ converges, and the theorem is proved.

8. Asymptotic and integral expressions for the J-fraction

We shall now consider a different approach to the problem of relating to the J-fraction the leading element $\rho_{11}(z)$ of a reciprocal of the J-matrix (4.5). We suppose still that the J-fraction satisfies the conditions (4.2), and shall use the ideas of §6, particularly Theorem 6.3, to obtain conditions under which the approximants $A_n(z)/B_n(z)$ approximate to $\rho_{11}(z)$ in the asymptotic sense of the following definition.

Definition 8.1. *Consider the domain*

$$S: \qquad \alpha \leqq \arg z \leqq \pi - \alpha \ (0 < \alpha < \pi/2), \qquad \mathfrak{J}(z) \geqq \delta > 0,$$

where α and δ are arbitrary positive numbers. A function $f(z)$ is said to be represented asymptotically by the J-fraction (4.1) *if*

$$\lim_{z=\infty} z^{2n}\left(f(z) - \frac{A_n(z)}{B_n(z)}\right) = 0, \qquad n = 1, 2, 3, \cdots, \tag{8.1}$$

as z approaches ∞ in S.

This is the same thing as to say that $f(z)$ is represented asymptotically by the series expansion of the J-fraction into powers of $1/z$.

Theorem 8.1. *A function $f(z)$ is represented asymptotically by the J-fraction* (4.1) *if and only if for every n there exists a number M_n such that the value of $f(z)$ is in the circle $K_n(z)$ for z in S and $\mathfrak{J}(z) > M_n$.*[16]

Proof. To prove that the condition is sufficient, let $w_q(z) = f(z)$, $q = 1, 2, 3, \cdots$, in (6.8), where $f(z)$ is any function satisfying the condition of the theorem. Then, from (6.8) we have:

$$z^{2p}\left(f(z) - \frac{A_p(z)}{B_p(z)}\right) = \frac{z^{2p}\rho_{p+1,p}(z)}{B_p(z)B_{p-1}(z)}, \qquad p = 1, 2, 3, \cdots. \tag{8.2}$$

If then $\mathfrak{J}(z) > M_{p+1}$, z in S, so that $f(z)$ has its value in $K_{p+1}(z)$, it follows from the formula (6.19) that

$$z^{2p}\left(f(z) - \frac{A_p(z)}{B_p(z)}\right) = \frac{H_p(z)}{\mathfrak{J}(z)},$$

where $H_p(z)$ is bounded in S; hence (8.1) holds.

We now suppose conversely that $f(z)$ is represented asymptotically by the J-fraction, and form the expression:

$$z^{2n+1}\left[\left(\frac{A_{n+1}(z)}{B_{n+1}(z)} - f(z)\right) + \left(f(z) - \frac{A_n(z)}{B_n(z)}\right)\right].$$

By the determinant formula, this is equal to $(a_1a_2 \cdots a_n)^2 + Q/z$, where Q is bounded in S. Hence we see that

$$\lim_{z=\infty} z^{2n+1}\left(f(z) - \frac{A_n(z)}{B_n(z)}\right) = (a_1a_2 \cdots a_n)^2, \tag{8.3}$$

[16] For the case where the b_p are real and $f(z)$ is analytic, R. Nevanlinna [12] has proved the same theorem with the stronger formulation that $M_n = 0$.

as z tends to ∞ in S. Let v_n be determined by the relation $f(z) = t_1 t_2 \cdots t_n(z; v_n)$; and recall that $f(z)$ is in the circle $K_n(z)$ if and only if $\mathfrak{I}(v_n) \leqq 0$. On substituting this value of $f(z)$ in (8.3) we obtain (cf. (2.7)):

$$\lim_{z=\infty} z^{2n+1}\left(\frac{A_n(z) - v_n A_{n-1}(z)}{B_n(z) - v_n B_{n-1}(z)} - \frac{A_n(z)}{B_n(z)}\right) = \lim_{z=\infty} z^{2n+1} \frac{v_n}{a_n B_n(z)[B_n(z) - v_n B_{n-1}(z)]}$$
$$= (a_1 a_2 \cdots a_n)^2.$$

Thus,

$$zv_n = \frac{za_n B_n(z)^2[(a_1 a_2 \cdots a_n)^2 + \epsilon(z)]}{z^{2n+1} + a_n B_n(z) B_{n-1}(z)[(a_1 a_2 \cdots a_n)^2 + \epsilon(z)]},$$

where $\epsilon(z)$ approaches 0 as z approaches ∞ in S. Therefore, since the coefficient of z^n in $B_n(z)$ is $(a_1 a_2 \cdots a_n)^{-1}$, zv_n converges to the positive limit a_n. Consequently, there exists a number M_n such that $\mathfrak{I}(v_n) \leqq 0$ and $f(z)$ is in $K_n(z)$, if $\mathfrak{I}(z) > M_n$ and z is in S.

Furthermore, Theorem 6.3 enables us to connect with the J-fractions "Stieltjes integrals" of the form (1.2) with range of integration extended over the whole real axis. Any function $f(z)$ which, for $\mathfrak{I}(z) > 0$, is analytic and has its values in the circles $K_n(z)$, $n = 1, 2, 3, \cdots$, will be called *equivalent* to the J-fraction. It is to be recalled that in the limit-point case, there is but one equivalent function, namely the value of the J-fraction; while in the limit-circle case there are infinitely many. If in (6.8) we suppose that $w_q(z) = f(z)$, $q = 1, 2, 3, \cdots$, then from (6.20) with $p = q = 1$ we obtain for any equivalent function the following estimate:

$$f(z) = \frac{1}{z} + \frac{g(z)}{z\mathfrak{I}(z)}, \quad \text{where} \quad |g(z)| \leqq C \quad \text{if} \quad \mathfrak{I}(z) > 0, \tag{8.4}$$

C being independent of z. From this we deduce the following general theorem:

THEOREM 8.2. *A function $f(z)$ has a Stieltjes integral representation of the form*

$$f(z) = \int_{-\infty}^{+\infty} \frac{d\phi(u)}{z+u}, \qquad \phi(+\infty) - \phi(-\infty) = 1, \tag{8.5}$$

in which $\phi(u)$ is real, bounded, and monotone nondecreasing, if $f(z)$ satisfies all of the following three conditions for $\mathfrak{I}(z) > 0$:

(i) *$f(z)$ is analytic*; (ii) $\mathfrak{I}[f(z)] \leqq 0$; (iii) The estimate (8.4) *holds*.*

PROOF. Suppose that $0 < y < c$, and consider the contour Γ in the upper half of the z-plane consisting of: the straight line segment from $A = -c^2 + iy$ to $B = c^2 + iy$, the straight line segment from B to $B' = c^2 + ic$, the arc of the circle with center at the origin through B' to $A' = -c^2 + ic$, and, finally, the

* The condition (iii) is not necessary, as can be seen from the example $\phi(u) = 0$ for $u \leqq 1, \phi(u) = 1 - u^{-(1/2)}$ for $u > 1$. This shows simultaneously that not every integral of the form (8.5) is equivalent to a J-fraction.

straight line segment from A' to A. Inasmuch as $f(z)$ is analytic in the domain interior to Γ we have, using (8.4), and evaluating explicitly $\int_{A'}^{B'} dz/z$:

$$\int_A^B f(z)\,dz = -\int_B^{B'} f(z)\,dz - \int_{B'}^{A'} f(z)\,dz - \int_{A'}^{A} f(z)\,dz$$

$$= -\pi i + 2i \arctan(y/c^2) + \frac{H}{c^2}\log\frac{y}{c} + \frac{H_1}{c},$$

where H and H_1 are bounded as c tends to ∞. Hence it follows that

$$\lim_{c=\infty}\int_{-c^2}^{+c^2} f(x+iy)\,dx = -\pi i, \quad \text{if} \quad y > 0. \tag{8.6}$$

Let $f(x+iy) = v(x, y) - iw(x, y)$, where $v(x, y)$ and $w(x, y)$ are real functions; then by (ii) $w(x, y) \geqq 0$ for $y > 0$. From (8.6) we conclude that

$$\lim_{c=\infty}\int_{-c^2}^{+c^2} w(x, y)\,dx = \pi. \tag{8.7}$$

Moreover, $\psi(u, y) = \int_0^u w(x, y)\,dx$ is a monotone nondecreasing function of u, is bounded by (8.7), and

$$\psi(+\infty, y) - \psi(-\infty, y) = \pi. \tag{8.8}$$

A well-known theorem[17] states that there exists a bounded monotone nondecreasing function $\psi(u)$ such that $\psi(+\infty) - \psi(-\infty) = \pi$, and a sequence $y_1, y_2, y_3, \cdots$ of positive numbers approaching the limit 0 such that

$$\lim_{n=\infty}\psi(u, y_n) = \psi(u) \tag{8.9}$$

at all points u where $\psi(u)$ is continuous.

If z is any point within Γ, then Cauchy's integral formula gives:

$$f(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f(s)\,ds}{s-z}.$$

Using (8.4) one may readily verify that for $c \to \infty$ this goes over into

$$f(z) = \frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{f(u+iy)\,du}{u+iy-z}, \tag{8.10}$$

where the integral is to be regarded in the sense of Cauchy's principal value. Let z^* be the point outside Γ which is symmetrical to the point z with respect to the line segment AB. Then we must have:

$$0 = \frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{f(u+iy)\,du}{u+iy-z^*} \quad \text{or} \quad 0 = \frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{\overline{f(u+iy)}\,du}{u-iy-\bar{z}^*}.$$

[17] For the proof see, for instance, Perron [14], pp. 394–395. This theorem has been applied in almost all investigations on problems of the kind considered here. The idea goes back to Stieltjes, and was developed and extended by Hilbert as one of the most important tools in his theory of infinite quadratic forms. See Hilbert [11] (book), p. 113 and 116.

Inasmuch as $u + iy - z = u - iy - \bar{z}^*$ we then have, on subtracting the last equation from the equation (8.10) and then introducing the function $\psi(u, y)$:

$$f(z) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{-w(u, y)\, du}{u + iy - z} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{d_u \psi(u, y)}{z - u - iy}.$$

On letting y approach 0 over the sequence y_n for which (8.9) holds, one then finds by a well-known argument (see footnote 17):

$$f(z) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{d\psi(u)}{z - u},$$

or, if we define $\phi(u)$ by $\phi(u) = -(1/\pi)\psi(-u)$, this becomes (8.5). Since $\phi(u)$ is bounded and monotone, the integral converges absolutely (not merely as Cauchy's principal value) and uniformly for z in any region at a positive distance from the real axis. The function $\phi(u)$ is given at all points of continuity by

$$\pi \cdot \phi(u) = \lim_{y=0} \int_0^{-u} \mathfrak{I}[f(x + iy)]\, dx, \tag{8.11}$$

where y approaches 0 over the sequence y_n. Since $f(z)$ is now expressed as an integral (8.5) the inversion process of Stieltjes[18] gives $\phi(u)$ in terms of $f(z)$, and shows, simultaneously, that (8.11) holds no matter how y approaches 0 over positive values. Thus $\phi(u)$ is determined uniquely by $f(z)$ to an additive constant at all points of continuity.

That the integrals ("moments")

$$\int_{-\infty}^{+\infty} u^p d\phi(u), \qquad p = 0, 1, 2, \cdots, \tag{8.12}$$

do not all exist when a single one of the b_p is nonreal may be argued from the fact that they are real if they exist, and from a theorem of H. Hamburger.[19]

The considerations of this section are closely connected with the "moment problem": To determine a real bounded nondecreasing function $\phi(u)$ taking on infinitely many different values and satisfying the infinite system of equations:

$$c_p = \int_{-\infty}^{+\infty} u^p d\phi(u), \qquad p = 0, 1, 2, \cdots, \tag{8.13}$$

where the c_p are given real numbers.

By the theorem of Hamburger just cited, $\phi(u)$ is a solution of (8.13) if and only if the function

$$f(z) = \int_{-\infty}^{+\infty} \frac{d\phi(u)}{z + u} \tag{8.14}$$

---

[18] Stieltjes [15], Chapter VI.

[19] Hamburger [5], Part I, Theorem IX (p. 268 ff).

is represented asymptotically by the power series $P(1/z)$ with the given numbers as coefficients:

$$f(z) \sim \frac{c_0}{z} - \frac{c_1}{z^2} + \frac{c_2}{z^3} - \cdots . \tag{8.15}$$

This is the same as saying that $f(z)$ is represented asymptotically by the J-fraction (4.1) "associated" (cf. Perron [14], §61) with $P(1/z)$. Hence, *all* the solutions of the moment problem are obtained by finding all functions $f(z)$ of the form (8.14) asymptotically equal to the J-fraction. By the theorem of Nevanlinna mentioned above and Theorem 8.2 these functions $f(z)$ are *just* those functions "equivalent" to the J-fraction. In this way one arrives at the complete solution of the moment problem.

It follows immediately that the moment problem is "determinate" in the limit-point case, and "indeterminate" in the limit-circle case.

NORTHWESTERN UNIVERSITY,
EVANSTON, ILLINOIS.

BIBLIOGRAPHY

1. RICHARD A. BETH, *Untersuchungen über die Spektraldarstellung von J-Formen*, Thesis, University of Frankfurt a.M., 1932, 69 pages.
2. T. CARLEMAN, *Sur les équations intégrales singulières à noyau symétrique*, Uppsala, 1923, pp. 189–220.
3. H. T. DAVIS, *The Theory of Linear Operators*, Bloomington, 1936, Chapters III and XII.
4. J. GROMMER, *Ganze transcendente Funktionen mit lauter reellen Nullstellen*, Journal für die reine und angewandte Mathematik, vol. 144 (1914), pp. 212–238.
5. H. HAMBURGER, *Über eine Erweiterung des Stieltjesschen Momentenproblems*, Parts I, II, and III, Mathematische Annalen, vol. 81 (1920), pp. 235–319; vol. 82 (1920) pp. 120–164; and pp. 168–187.
6. E. HELLINGER, *Zur Stieltjesschen Kettenbruchtheorie*, Mathematische Annalen, vol. 86 (1922), pp. 18–29.
7. ———, *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen*, Journal für die reine und angewandte Mathematik, vol. 136 (1909), pp. 210–271.
8. E. HELLINGER AND O. TOEPLITZ, *Zur Einordnung der Kettenbruchtheorie in die Theorie der quadratischen Formen von unendlichvielen Veränderlichen*, Journal für die reine und angewandte Mathematik, vol. 144 (1914), pp. 212–238.
9. ———, *Grundlagen für eine Theorie der unendlichen Matrizen*, Mathematische Annalen, vol. 69 (1910), pp. 289–330.
10. ———, *Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten*, Encyklopädie der mathematischen Wissenschaften, II C 13, No. 18 and No. 43.
11. D. HILBERT, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, Vierte Mitteil., Nachrichten der königlichen Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse, 1906, p. 219 ff; or book of same title, Leipzig, 1912, Chapter IV.
12. R. NEVANLINNA, *Asymptotische Entwicklungen beschränkter Funktionen und das Stieltjessche Momentenproblem*, Annales academiae Fennicae, Series A, vol. 18 (1922), No. 5.
13. J. F. PAYDON AND H. S. WALL, *The continued fraction as a sequence of linear transformations*, Duke Mathematical Journal, vol. 9 (1942), pp. 360–372.

14. O. PERRON, *Die Lehre von den Kettenbrüchen*, 2nd ed., Leipzig and Berlin, 1929.

15. T. J. STIELTJES, *Recherches sur les fractions continues*, Oeuvres, vol. II, pp. 402–566. Published also in vols. 8 and 9 of Annales de Toulouse; and in vol. 32 of Mémoires présentés à l'Académie des Sciences de l'Institut National de France.

16. E. B. VAN VLECK, *On an extension of the 1894 memoir of Stieltjes*, Transactions of the American Mathematical Society, vol. 4 (1903), pp. 297–332.

17. ———, *On the convergence of continued fractions with complex elements*, Transactions of the American Mathematical Society, vol. 2 (1901), pp. 215–233.

18. H. WEYL, *Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen*, Mathematische Annalen, vol. 68 (1910), pp. 220–269.

19. ———, *Über das Pick-Nevanlinna'sche Interpolationsproblem und sein infinitesimales Analogon*, Annals of Mathematics, Series 2, vol. 36 (1935), pp. 230–254.

REMARKS ON TWO-LEAVED ORIENTABLE COVERING MANIFOLDS OF CLOSED MANIFOLDS

By Tsai-han Kiang

(Received October 5, 1942)

1. In the present note the covering complexes considered are *without branch point*, and the covering manifolds are, in addition, *of finite number of leaves.*[1] Our purpose is to establish the following two theorems.

Theorem 1. *Any orientable covering manifold of a closed non-orientable n-manifold M is a covering manifold of the 2-leaved orientable covering manifold*[2] *of M, and hence a covering manifold of M of even number of leaves.*

Theorem 2. *A necessary and sufficient condition for the existence of a simplicial topological self-mapping of a closed orientable n-manifold $\bar{M}$, which is involutory, without fixed point and orientation-preserving (orientation-reversing),*[3] *is that $\bar{M}$ be a 2-leaved covering manifold of a closed orientable (nonorientable) n-manifold M.*

2. Theorem 1 follows at once from the following two lemmas.

Lemma 1. *Let K be a connected n-complex and F its fundamental group. Suppose that G and H are two subgroups of F and that H is a subgroup of G. If K^* and K^{**} are the covering complexes of K, determined by G and H respectively,*[4] *then K^{**} is a covering complex of K^*.*[5]

Proof. From hypothesis the fundamental groups F^* and F^{**} of K^* and K^{**} are simply isomorphic to G and H respectively.[4] By virtue of the simple isomorphism between F^* and G, there is a definite subgroup of F^*, which is simply isomorphic to F^{**}. Let $\bar{K}$ be the covering complex of K^*, determined by this subgroup of the fundamental group F^* of K^*. Then $\bar{K}$ is a covering complex of K.[6] Since the fundamental group of $\bar{K}$ is simply isomorphic to F^{**} and therefore to H, and since a covering complex of K is uniquely determined by a subgroup of the fundamental group of K,[4] $\bar{K}$ and K^{**} are not distinct.[7] Hence K^{**} is a covering complex of K^*.

Lemma 2. *Let $\bar{M}$ be a covering manifold*[8] *of a closed n-manifold M with a cellular decomposition M_a,*[9] *and $\bar{M}_a$ the cellular decomposition of $\bar{M}$ derived from*

[1] The topological terms will be defined as in Seifert-Threlfall, *Lehrbuch der Topologie* (1934). This book will be referred to hereafter as ST.

[2] Cf. ST, p. 272.

[3] ST, p. 129.

[4] ST, p. 193.

[5] Cf. K. Reidemeister, *Einführung in die Kombinatorische Topologie* (1932), p. 125.

[6] ST, p. 194.

[7] ST, pp. 182–183.

[8] A finite covering complex of a closed n-manifold is obviously a closed n-manifold.

[9] ST, p. 341.

M_a.[10] *If an edge-path $\bar{U}$ on $\bar{M}_a$ at a vertex $\bar{O}_1$ is closed, then $\bar{U}$ and its closed ground-path U on M_a at the ground-vertex O of $\bar{O}_1$ are either both orientation-preserving*[11] *or both not.*[12]

PROOF. Let M_b be the cellular decomposition of M dual to M_a. Dual to the vertices and edges of the closed U, in the order as they appear alternately in U, there is on M_b the closed sequence V of incident cells of dimensions n and $n - 1$:

$$V = C_0^n C_0^{n-1} C_1^n C_1^{n-1} \cdots C_t^n C_t^{n-1},$$

where C_0^n is the dual of the vertex O. Let $\bar{M}_b$ denote the cellular decomposition of $\bar{M}$ derived from M_b of M. Then $\bar{M}_b$ is dual to $\bar{M}_a$, and the dual $\bar{V}$ on $\bar{M}_b$ of $\bar{U}$

$$\bar{V} = \bar{C}_0^n \bar{C}_0^{n-1} \bar{C}_1^n \bar{C}_1^{n-1} \cdots \bar{C}_t^n \bar{C}_t^{n-1}$$

covers V in the sense that $\bar{C}_i^n$, $\bar{C}_i^{n-1}$ covers C_i^n, C_i^{n-1}. Since the orientations of $\bar{C}_i$ can be derived from those of C_i, our lemma follows at once.

PROOF OF THEOREM 1. Let the n-complex in Lemma 1 be a closed non-orientable n-manifold M with a definite cellular decomposition. Its only 2-leaved orientable covering manifold M^* is determined by the subgroup G of index 2 of the fundamental group of M, whose elements are the classes of homotopically deformable closed orientation-preserving edge-paths of M at a vertex O^2. A necessary and sufficient condition that a manifold be orientable is obviously that all the closed edge-paths on M at a vertex are orientation-preserving. Suppose that a covering manifold M^{**} of M is orientable. From Lemma 2, its fundamental group is simply isomorphic to a subgroup of G. From Lemma 1, it is a covering manifold of M^*, and hence a covering manifold of M of even number of leaves.

From Lemma 2 and the proof of our theorem, we have the following immediate consequences:

COROLLARY 1. *Any covering manifold of a closed orientable manifold is orientable.*

COROLLARY 2. *Let M be a closed nonorientable manifold and G the group of all classes of homotopically deformable closed orientation-preserving paths on M at a point O. A covering manifold of M, determined by a subgroup of H of the fundamental group of M at O, is orientable when and only when H is a subgroup of G.*

3. PROOF OF THEOREM 2. SUFFICIENCY. Suppose that M is a closed n-manifold, and that $\bar{M}$ a 2-leaved, and therefore regular, orientable covering manifold of M. The covering motion (Deckbewegung) f on $\bar{M}$ is a simplicial

---

[10] ST, p. 189, p. 272.

[11] ST, p. 191. Notice that an orientation-preserving or orientation-reversing edge-path may have double points.

[12] This lemma is tacitly used in ST, p. 272, in discussion of special M and $\bar{M}$.

topological self-mapping of $\bar{M}$, which is involutory and without fixed point. It remains to show, as follows, that f is orientation-preserving or orientation-reversing according as M is orientable or nonorientable.

Let M_a, m_b, $\bar{M}_a$, $\bar{M}_b$ have the same meaning as in Lemma 2 and its proof, but let M_a be simplicial. Suppose that $\bar{M}$ is determined by the subgroup H of the fundamental group F of M_a with reference to a vertex O of M_a as the initial point of closed edge-paths. Denote by $\bar{O}_1$ and $\bar{O}_2$ the two covering vertices of O. Take an arbitrary closed edge-path U on M_a at O. Denote its covering path on $\bar{M}_a$ at $\bar{O}_1$ by $\bar{U}$. $\bar{U}$ begins at $\bar{O}_1$ and ends at $\bar{O}_2$.

Now let the n-cell on M_b dual to O be C^n, and the n-cells on $\bar{M}_b$ dual to $\bar{O}_i$, $i = 1, 2$, be $\bar{C}_i^n$. The continuation of orientation along $\bar{U}$ can be derived from that along that along U.[13] When M is orientable (nonorientable), U is orientation-preserving (orientation-reserving). Hence the orientations of $\bar{C}_1^n$ derived from the same orientation (opposite orientations) of C^n are coherent on $\bar{M}_b$. Since f maps $\bar{C}_1^n$ and $\bar{C}_2^n$ onto one another, and their orientations derived from the same orientation of C^n, f is orientation-preserving (orientation-reversing) on $\bar{M}$.

Necessity. Suppose that $\bar{M}$ is a closed orientable n-manifold on which there is a simplicial topological involutory self-mapping f without fixed point. Through identification of pairs of corresponding points on $\bar{M}$ under f, there results a space M. Since the mapping of $\bar{M}$ on M, defined by the identification, is continuous, M is connected. Since f is simplicial and topological in the small, M is a complex and a closed n-manifold respectively. Then obviously $\bar{M}$ fulfills the condition of being a 2-leaved covering manifold of M.

Finally, by virtue of the result in the proof of sufficiency, M is orientable or nonorientable according as f is orientation-preserving or orientation-reversing.

From the fact that the Euler-Poincaré characteristics of M is half that of $\bar{M}$, and from the Poincaré duality theorem for orientable manifold, we have for $n = 2m$ and for $n = 2$ and orientable M the following respectively:

Corollary 3. *If on a closed orientable* $(2m)$*-manifold there exists a simplicial topological involutory self-mapping without fixed point, the* m^{th} *Betti number of the manifold must be even.*

Corollary 4. *On a closed orientable* 2*-manifold of even genus there is no simplicial topological orientation-preserving self-mapping, which is involutory and without fixed point.*[14]

The National University of Peking
and Academia Sinica

[13] ST, p. 271.

[14] Cf. A Komatu, *Uber die dreidimensionalen nichtorientierbaren Mannigfaltigkeiten*, Satz 2, Proc. Phys.-Math. Soc. Japan, Vol. 18 (1936), pp. 135–141.